PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
A01H 5/00, 5/10, A23D 7/00, A23K 1/14, C07H 21/04, C07K 4/10, C12N 15/04, 15/63, 15/82

A1

(11) International Publication Number: **WO 99/02030**

(43) International Publication Date: 21 January 1999 (21.01.99)

(21) International Application Number: PCT/US98/14528

(22) International Filing Date: 10 July 1998 (10.07.98)

(30) Priority Data:
60/052,249 11 July 1997 (11.07.97) US

(71) Applicant *(for all designated States except US)*: WASHINGTON STATE UNIVERSITY RESEARCH FOUNDATION [US/US]; N.E. 1615 Eastgate Boulevard, Pullman, WA 99164–1802 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: BOHLMANN, Joerg [DE/DE]; Sophienstrasse 21, D–0745 Jena (DE). STEELE, Christopher, L. [US/US]; 617 Campbell Street, Ardmore, OK 973401 (US). CROTEAU, Rodney, B. [US/US]; N.E. 1835 Valley Road, Pullman, WA 99163 (US).

(74) Agent: SHELTON, Dennis, K.; Christensen O'Connor Johnson & Kindness PLLC, Suite 2800, 1420 Fifth Avenue, Seattle, WA 98101 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: MONOTERPENE SYNTHASES FROM GRAND FIR (*ABIES GRANDIS*)

(57) Abstract

cDNAs encoding myrcene synthase, (–)–limonene synthase and (–)–pinene synthase from Grand fir (*Abies grandis*) haven been isolated and sequenced, and the corresponding amino acid sequences have been determined. Accordingly, isolated DNA sequences (SEQ ID NO:1; SEQ ID NO:3 and SEQ ID NO:5) are provided which code for the expression of myrcene synthase (SEQ ID NO:2), (–)–pinene synthase (SEQ ID NO:4) and (–)–limonene synthase (SEQ ID NO:6) , respectively, from Grand fir (*Abies grandis*). In other aspects, replicable recombinant cloning vehicles are provided which code for myrcene synthase, (–)–limonene synthase and (–)–pinene synthase, or for a base sequence sufficiently complementary to at least a portion of myrcene synthase, (–)–limonene synthase or (–)–pinene synthase DNA or RNA to enable hybridization therewith. In yet other aspects, modified host cells are provided that have been transformed, transfected, infected and/or injected with a recombinant cloning vehicle and/or DNA sequence encoding myrcene synthase, (–)–limonene synthase or (–)–pinene synthase. Thus, systems and methods are provided for the recombinant expression of the aforementioned recombinant monoterpene synthases that may be used to facilitate their production, isolation and purification in significant amounts. Recombinant myrcene synthase, (–)–limonene synthase or (–)–pinene synthase may be used to obtain expression or enhanced expression of myrcene synthase, (–)–limonene synthase or (–)–pinene synthase in plants in order to enhance the production of monoterpenoids, or may be otherwise employed for the regulation or expression of myrcene synthase, (–)–limonene synthase and (–)–pinene synthase, or the production of their products.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

MONOTERPENE SYNTHASES FROM GRAND FIR (*Abies grandis*)

Field of the Invention

The present invention relates to nucleic acid sequences which code for monoterpene synthases from gymnosperm plant species, in particular from Grand fir (*Abies grandis*), including limonene synthase, myrcene synthase, and pinene synthase, to vectors containing the sequences, to host cells containing the sequences, to plant seeds expressing the sequences and to methods of producing recombinant monoterpene synthases and their mutants.

Background of the Invention

Chemical defense of conifer trees against bark beetles and their associated fungal pathogens relies primarily upon constitutive and inducible oleoresin biosynthesis (Johnson, M.A., and Croteau, R. (1987) in *Ecology and Metabolism of Plant Lipids* (Fuller, G., and Nes, W.D., eds.) pp. 76-91, American Chemical Society Symposium Series 325, Washington, DC; Gijzen, M., Lewinsohn, E., Savage, T.J., and Croteau, R.B. (1993) in *Bioactive Volatile Compounds from Plants* (Teranishi, R., Buttery, R.G., and Sugisawa, H., eds.) pp. 8-22, American Chemical Society Symposium Series 525, Washington, DC). This defensive secretion is a complex mixture of monoterpene and sesquiterpene olefins (turpentine) and diterpene resin acids (rosin) that is synthesized constitutively in the epithelial cells of specialized structures, such as resin ducts and blisters or, in the case of induced oleoresin formation, in undifferentiated cells surrounding wound sites (Lewinsohn, E., Gijzen, M., Savage, T.J., and Croteau, R. (1991) *Plant Physiol.* **96**:38-43). The volatile fraction of conifer oleoresin, which is toxic to both bark beetles and their fungal

associates (Raffa, K.F., Berryman, A.A., Simasko, J., Teal, W., and Wong, B.L. (1985) *Environ. Entomol.* **14**:552-556), may consist of up to 30 different monoterpenes (Lewinsohn, E., Savage, T.J., Gijzen, M., and Croteau, R. (1993) *Phytochem. Anal.* **4**:220-225), including acyclic types (e.g., myrcene), monocyclic types (e.g., limonene) and bicyclic types (e.g., pinenes) (FIGURE 1). Although the oleoresin is toxic, many bark beetle species nevertheless employ turpentine volatiles in host selection and can convert various monoterpene components into aggregation or sex pheromones to promote coordinated mass attack of the host (Gijzen, M., Lewinsohn, E., Savage, T.J., and Croteau, R.B. (1993) in *Bioactive Volatile Compounds from Plants* (Teranishi, R., Buttery, R.G., and Sugisawa, H., eds.) pp. 8-22, American Chemical Society Symposium Series 525, Washington, DC; Byers, J.A. (1995) in *Chemical Ecology of Insects 2* (Cardé, R.T., and Bell, W.J., eds.) pp. 154-213, Chapman and Hall, New York). In Grand fir (*Abies grandis*), increased formation of oleoresin monoterpenes, sesquiterpenes and diterpenes is induced by bark beetle attack (Lewinsohn, E., Gijzen, M., Savage, T.J., and Croteau, R. (1991) *Plant Physiol.* **96**:38-43; Raffa, K.F., and Berryman, A.A. (1982) *Can. Entomol.* **114**:797-810; Lewinsohn, E., Gijzen, M., and Croteau, R. (1991) *Plant Physiol.* **96**:44-49), and this inducible defense response is mimicked by mechanically wounding sapling stems (Lewinsohn, E., Gijzen, M., Savage, T.J., and Croteau, R. (1991) *Plant Physiol.* **96**:38-43; Lewinsohn, E., Gijzen, M., and Croteau, R. (1991) *Plant Physiol.* **96**:44-49; Funk, C., Lewinsohn, E., Stofer Vogel, B., Steele C., and Croteau, R. (1994) *Plant Physiol.* **106**:999-1005). Therefore, Grand fir has been developed as a model system to study the biochemical and molecular genetic regulation of constitutive and inducible terpene biosynthesis in conifers (Steele, C., Lewinsohn, E., and Croteau, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**:4164-4168).

Most monoterpenes are derived from geranyl diphosphate, the ubiquitous C_{10} intermediate of the isoprenoid pathway, by synthases which catalyze the divalent metal ion-dependent ionization (to 1, FIGURE 1) and isomerization of this substrate to enzyme-bound linalyl diphosphate which, following rotation about C2-C3, undergoes a second ionization (to 2, FIGURE 1) followed by cyclization to the α-terpinyl cation, the first cyclic intermediate en route to both monocyclic and bicyclic products (Croteau, R., and Cane, D.E. (1985) *Methods Enzymol.* **110**:383-405; Croteau, R. (1987) *Chem. Rev.* **87**:929-954) (FIGURE 1). Acyclic monoterpenes, such as myrcene, may arise by deprotonation of carbocations 1 or 2, whereas the isomerization step to linalyl diphosphate is required in the case of cyclic

types, such as limonene and pinenes, which cannot be derived from geranyl diphosphate directly because of the geometric impediment of the *trans*-double bond at C2-C3 (Croteau, R., and Cane, D.E. (1985) *Methods Enzymol.* **110**:383-405; Croteau, R. (1987) *Chem. Rev.* **87**:929-954). Many monoterpene synthases catalyze the formation of multiple products, including acyclic, monocyclic and bicyclic types, by variations on this basic mechanism (Gambliel, H., and Croteau, R. (1984) *J. Biol. Chem.* **259**:740-748; Croteau, R., Satterwhite, D.M., Cane, D.E., and Chang, C.C. (1988) *J. Biol. Chem.* **263**:10063-10071; Croteau, R., and Satterwhite, D.M. (1989) *J. Biol. Chem.* **264**:15309-15315). For example, (-)-limonene synthase, the principal monoterpene synthase of spearmint (*Mentha spicata*) and peppermint (*M.* x *piperita*), produces small amounts of myrcene, (-)-α-pinene and (-)-β-pinene in addition to the monocyclic product (Rajaonarivony, J.I.M., Gershenzon, J., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **296**:49-57; Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol Chem.* **268**:23016-23024. Conversely, six different inducible monoterpene synthase activities have been demonstrated in extracts of wounded Grand fir stem (Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**:267-273) indicating that formation of acyclic, monocyclic and bicyclic monoterpenes in this species involves several genes encoding distinct catalysts. The inducible (-)-pinene synthase has been purified (Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173), and isotopically sensitive branching experiments employed to demonstrate that this enzyme synthesizes both (-)-α- and (-)-β-pinene (Wagschal, K., Savage, T.J., and Croteau, R. (1991) *Tetrahedron* **47**:5933-5944).

Deciphering the molecular genetic control of oleoresinosis and examining structure-function relationships among the monoterpene synthases of Grand fir requires isolation of the cDNA species encoding these key enzymes. Although a protein-based cloning strategy was recently employed to acquire a cDNA for the major wound-inducible diterpene synthase from Grand fir, abietadiene synthase (Funk, C., Lewinsohn, E., Stofer Vogel, B., Steele C., and Croteau, R. (1994) *Plant Physiol.* **106**:999-1005; LaFever, R.E., Stofer Vogel, B., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **313**:139-149; Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268), all attempts at the reverse genetic approach to cloning of Grand fir monoterpene synthases have failed (Steele, C., Lewinsohn, E., and Croteau, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**:4164-4168). As an alternative, a similarity-based PCR strategy was developed (Steele, C.,

Lewinsohn, E., and Croteau, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**:4164-4168) that employed sequence information from terpene synthases of angiosperm origin, namely a monoterpene synthase, (-)-4*S*-limonene synthase, from spearmint (*Mentha spicata*, Lamiaceae) (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024), a sesquiterpene synthase, 5-*epi*-aristolochene synthase, from tobacco (*Nicotiana tabacum*, Solanaceae) (Facchini, P.J., and Chappell, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**:11088-11092), and a diterpene synthase, casbene synthase, from castor bean (*Ricinus communis*, Euphorbiaceae) (Mau, C.J.D., and West, C.A. (1994) *Proc Natl. Acad. Sci. USA* **91**:8497-8501).

Monoterpenes have significant potential for cancer prevention and treatment. Monoterpenes such as limonene, perillyl alcohol, carvone, geraniol and farnesol not only reduce tumor incidence and slow tumor proliferation, but have also been reported to cause regression of established solid tumors by initiating apoptosis (Mills J.J., Chari R.S., Boyer I.J., Gould M.N., Jirtle R.L., *Cancer Res.*, **55**:979-983, 1995). Terpenes have activity against cancers such as mammary, colon, and prostate. Clinical trials are being pursued (Seachrist L, *J. NIH Res.* **8**:43) in patients with various types of advanced cancers to validate the health benefits of dietary terpenes for humans. However, terpenes are present in Western diets at levels that are probably inadequate for any significant preventive health benefits. Daily supplementation of the diet with a terpene concentrate (10-20 g/day) would appear to be the most rational strategy for dietary therapy of diagnosed cases of cancer. This invention envisages the production of such nutritionally beneficial terpenes in vegetable oils consumed daily via the engineering of relevant genes from Grand fir into oil seed crop plants such as oil seed brassica (canola), soybean and corn.

Brief Description of the Drawings

The foregoing aspects and many of the attendant advantages of this invention will become more readily appreciated as the same becomes better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings, wherein:

FIGURE 1 is a schematic representation depicting the mechanism for the conversion of geranyl diphosphate to myrcene, (-)-limonene, β-phellandrene, (-)-α-pinene and (-)-β-pinene by monoterpene synthases from Grand fir. Formation of the monocyclic and bicyclic products requires preliminary isomerization of geranyl diphosphate to linalyl diphosphate. The acyclic product could be formed

from either geranyl diphosphate or linalyl diphosphate via carbocations 1 or 2. *OPP* denotes the diphosphate moiety.

FIGURE 2 is a sequence comparison of plant terpene synthases. A three-letter designation (*Tps*) for the gene family is proposed with sub-groups (*Tpsa* through *Tpsf*) defined by a minimum of 40 % amino acid identity between members. The numbers in parenthesis are the references to the published sequences.

FIGURE 3 depicts a GLC-MS analysis of the products of the recombinant protein encoded by *AG2.2* (SEQ ID NO:1), the sequence of the protein encoded by clone *AG2.2* being set forth in SEQ ID NO:2. The GLC profile of the total pentane-soluble products generated from geranyl diphosphate when incubated with a cell-free extract of *E. coli* XL1-Blue/pGAG2.2 is illustrated (**a**), as are the mass fragmentation patterns for the monoterpene product with R_t = 12.22 min (**b**) and for authentic myrcene (**c**).

FIGURE 4 depicts a GLC-MS analysis of the products of the recombinant protein encoded by *AG3.18* (SEQ ID NO:3), the sequence of the protein encoded by clone *AG3.18* (SEQ ID NO:3) being set forth in SEQ ID NO:4. The GLC profile of the total pentane-soluble products generated from geranyl diphosphate when incubated with a cell-free extract of *E. coli* XL1-Blue/pGAG3.18 is illustrated (**a**), as are the mass fragmentation patterns (selected ion mode) for the monoterpene products with R_t = 11.34 min (**b**), and R_t = 13.37 min (**d**), and for authentic α-pinene (**c**) and authentic β-pinene (**e**).

FIGURE 5 depicts a GLC-MS analysis of the products of the recombinant protein encoded by *AG10* (SEQ ID NO:5), the sequence of the protein encoded by clone *AG10* (SEQ ID NO:5) being set forth in SEQ ID NO:6. The GLC profile of the total pentane-soluble products generated from geranyl diphosphate when incubated with a cell-free extract of *E. coli* BL21(DE3)/pSBAG10 is illustrated (**a**), as are the mass fragmentation patterns for the principal monoterpene product with R_t = 13.93 min (**b**) and for authentic limonene (**c**).

Summary of the Invention

In accordance with the foregoing, cDNAs encoding myrcene synthase, (-)-limonene synthase and (-)-pinene synthase from Grand fir (*Abies grandis*) have been isolated and sequenced, and the corresponding amino acid sequences have been deduced. Accordingly, the present invention relates to isolated DNA sequences which code for the expression of myrcene synthase, such as the sequence designated SEQ ID NO:1 which encodes myrcene synthase from Grand fir (*Abies grandis*), for

the expression of (-)-pinene synthase, such as the sequence designated SEQ ID NO:3, which encodes the (-)-pinene synthase from Grand fir (*Abies grandis*) and for the expression of (-)-limonene synthase, such as the sequence designated SEQ ID NO:5, which encodes (-)-limonene synthase from Grand fir (*Abies grandis*). In other aspects, the present invention is directed to replicable recombinant cloning vehicles comprising a nucleic acid sequence, e.g., a DNA sequence which codes for a myrcene synthase, (-)-limonene synthase or (-)-pinene synthase, or for a base sequence sufficiently complementary to at least a portion of DNA or RNA encoding myrcene synthase, (-)-limonene synthase or (-)-pinene synthase to enable hybridization therewith (e.g., antisense RNA or fragments of DNA complementary to a portion of DNA or RNA molecules encoding myrcene synthase, (-)-limonene synthase or (-)-pinene synthase which are useful as polymerase chain reaction primers or as probes for any of the foregoing synthases or related genes). In yet other aspects of the invention, modified host cells are provided that have been transformed, transfected, infected and/or injected with a recombinant cloning vehicle and/or DNA sequence of the invention. Thus, the present invention provides for the recombinant expression of myrcene synthase, (-)-limonene synthase and (-)-pinene synthase, and the inventive concepts may be used to facilitate the production, isolation and purification of significant quantities of recombinant myrcene synthase, (-)-limonene synthase and (-)-pinene synthase (or of their primary enzyme products) for subsequent use, to obtain expression or enhanced expression of myrcene synthase, (-)-limonene synthase and (-)-pinene synthase in plants, microorganisms or animals, or may be otherwise employed in an environment where the regulation or expression of myrcene synthase, (-)-limonene synthase and (-)-pinene synthase is desired for the production of these synthases, or their enzyme products, or derivatives thereof. In another aspect, the present invention relates to manipulation of monoterpene production to enhance resistance to insects and/or accumulate nutritionally beneficial monoterpenes in oil seeds.

Detailed Description of the Preferred Embodiment

As used herein, the terms "amino acid" and "amino acids" refer to all naturally occurring L-α-amino acids or their residues. The amino acids are identified by either the single-letter or three-letter designations:

Asp	D	aspartic acid	Ile	I	isoleucine
Thr	T	threonine	Leu	L	leucine
Ser	S	serine	Tyr	Y	tyrosine

Glu	E	glutamic acid	Phe	F	phenylalanine
Pro	P	proline	His	H	histidine
Gly	G	glycine	Lys	K	lysine
Ala	A	alanine	Arg	R	arginine
Cys	C	cysteine	Trp	W	tryptophan
Val	V	valine	Gln	Q	glutamine
Met	M	methionine	Asn	N	asparagine

As used herein, the term "nucleotide" means a monomeric unit of DNA or RNA containing a sugar moiety (pentose), a phosphate and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of pentose) and that combination of base and sugar is called a nucleoside. The base characterizes the nucleotide with the four bases of DNA being adenine ("A"), guanine ("G"), cytosine ("C") and thymine ("T"). Inosine ("I") is a synthetic base that can be used to substitute for any of the four, naturally-occurring bases (A, C, G or T). The four RNA bases are A,G,C and uracil ("U"). The nucleotide sequences described herein comprise a linear array of nucleotides connected by phosphodiester bonds between the 3' and 5' carbons of adjacent pentoses.

The term "percent identity" means the percentage of amino acids or nucleotides that occupy the same relative position when two amino acid sequences, or two nucleic acid sequences are aligned side by side.

The term "percent similarity" is a statistical measure of the degree of relatedness of two compared protein sequences. The percent similarity is calculated by a computer program that assigns a numerical value to each compared pair of amino acids based on chemical similarity (e.g., whether the compared amino acids are acidic, basic, hydrophobic, aromatic, etc.) and/or evolutionary distance as measured by the minimum number of base pair changes that would be required to convert a codon encoding one member of a pair of compared amino acids to a codon encoding the other member of the pair. Calculations are made after a best fit alignment of the two sequences have been made empirically by iterative comparison of all possible alignments. (Henikoff, S. and Henikoff, J.G., *Proc. Nat'l. Acad. Sci. USA* **89**:10915-10919, 1992).

"Oligonucleotide" refers to short length single or double stranded sequences of deoxyribonucleotides linked via phosphodiester bonds. The oligonucleotides are chemically synthesized by known methods and purified, for example, on polyacrylamide gels.

The term "myrcene synthase" is used herein to mean an enzyme capable of generating multiple monoterpenes from geranyl diphosphate. The principal and characteristic monoterpene synthesized by myrcene synthase is myrcene, which constitutes at least about 60% of the monoterpene mixture synthesized by myrcene synthase from geranyl diphosphate.

The term "(-)-limonene synthase" is used herein to mean an enzyme capable of generating multiple monoterpenes from geranyl diphosphate. The principal and characteristic monoterpene synthesized by (-)-limonene synthase is (-)-limonene, which constitutes at least about 60% of the monoterpene mixture synthesized by (-)-limonene synthase from geranyl diphosphate.

The term "(-)-pinene synthase" is used herein to mean an enzyme capable of generating multiple monoterpenes from geranyl diphosphate. The principal and characteristic monoterpene synthesized by (-)-pinene synthase is (-)-pinene, which comprises at least about 60% of the monoterpene mixture synthesized by (-)-pinene synthase from geranyl diphosphate.

The abbreviation "SSPE" refers to a buffer used in nucleic acid hybridization solutions. The 20X (twenty times concentrate) stock SSPE buffer solution is prepared as follows: dissolve 175.3 grams of NaCl, 27.6 grams of $NaH_2PO_4H_2O$ and 7.4 grams of EDTA in 800 millilitres of H_2O. Adjust the pH to pH 7.4 with NaOH. Adjust the volume to one liter with H_2O.

The terms "alteration", "amino acid sequence alteration", "variant" and "amino acid sequence variant" refer to monoterpene synthase molecules with some differences in their amino acid sequences as compared to the corresponding, native, i.e., naturally-occurring, monoterpene synthases. Ordinarily, the variants will possess at least about 70% homology with the corresponding native monoterpene synthases, and preferably, they will be at least about 80% homologous with the corresponding, native monoterpene synthases. The amino acid sequence variants of the monoterpene synthases falling within this invention possess substitutions, deletions, and/or insertions at certain positions. Sequence variants of monoterpene synthases may be used to attain desired enhanced or reduced enzymatic activity, modified regiochemistry or stereochemistry, or altered substrate utilization or product distribution.

Substitutional monoterpene synthase variants are those that have at least one amino acid residue in the native monoterpene synthase sequence removed and a different amino acid inserted in its place at the same position. The substitutions may

be single, where only one amino acid in the molecule has been substituted, or they may be multiple, where two or more amino acids have been substituted in the same molecule. Substantial changes in the activity of the monoterpene synthase molecules of the present invention may be obtained by substituting an amino acid with a side chain that is significantly different in charge and/or structure from that of the native amino acid. This type of substitution would be expected to affect the structure of the polypeptide backbone and/or the charge or hydrophobicity of the molecule in the area of the substitution.

Moderate changes in the activity of the monoterpene synthase molecules of the present invention would be expected by substituting an amino acid with a side chain that is similar in charge and/or structure to that of the native molecule. This type of substitution, referred to as a conservative substitution, would not be expected to substantially alter either the structure of the polypeptide backbone or the charge or hydrophobicity of the molecule in the area of the substitution.

Insertional monoterpene synthase variants are those with one or more amino acids inserted immediately adjacent to an amino acid at a particular position in the native monoterpene synthase molecule. Immediately adjacent to an amino acid means connected to either the α-carboxy or α-amino functional group of the amino acid. The insertion may be one or more amino acids. Ordinarily, the insertion will consist of one or two conservative amino acids. Amino acids similar in charge and/or structure to the amino acids adjacent to the site of insertion are defined as conservative. Alternatively, this invention includes insertion of an amino acid with a charge and/or structure that is substantially different from the amino acids adjacent to the site of insertion.

Deletional variants are those where one or more amino acids in the native monoterpene synthase molecules have been removed. Ordinarily, deletional variants will have one or two amino acids deleted in a particular region of the monoterpene synthase molecule.

The terms "biological activity", "biologically active", "activity" and "active" refer to the ability of the monoterpene synthases of the present invention to convert geranyl diphosphate to a group of monoterpenes, of which myrcene is the principal and characteristic monoterpene synthesized by myrcene synthase, (-)-limonene is the principal and characteristic monoterpene synthesized by (-)-limonene synthase and (-)-pinene is the principal and characteristic monoterpene synthesized by (-)-pinene synthase. The monoterpenes produced by the monoterpene synthases of the present

invention are as measured in an enzyme activity assay, such as the assay described in Example 3. Amino acid sequence variants of the terpene synthases of the present invention may have desirable altered biological activity including, for example, altered reaction kinetics, substrate utilization product distribution or other characteristics such as regiochemistry and stereochemistry.

The terms "DNA sequence encoding", "DNA encoding" and "nucleic acid encoding" refer to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the translated polypeptide chain. The DNA sequence thus codes for the amino acid sequence.

The terms "replicable expression vector" and "expression vector" refer to a piece of DNA, usually double-stranded, which may have inserted into it a piece of foreign DNA. Foreign DNA is defined as heterologous DNA, which is DNA not naturally found in the host. The vector is used to transport the foreign or heterologous DNA into a suitable host cell. Once in the host cell, the vector can replicate independently of or coincidental with the host chromosomal DNA, and several copies of the vector and its inserted (foreign) DNA may be generated. In addition, the vector contains the necessary elements that permit translating the foreign DNA into a polypeptide. Many molecules of the polypeptide encoded by the foreign DNA can thus be rapidly synthesized.

The terms "transformed host cell," "transformed" and "transformation" refer to the introduction of DNA into a cell. The cell is termed a "host cell", and it may be a prokaryotic or a eukaryotic cell. Typical prokaryotic host cells include various strains of *E. coli*. Typical eukaryotic host cells are plant cells, such as maize cells, yeast cells, insect cells or animal cells. The introduced DNA is usually in the form of a vector containing an inserted piece of DNA. The introduced DNA sequence may be from the same species as the host cell or from a different species from the host cell, or it may be a hybrid DNA sequence, containing some foreign DNA and some DNA derived from the host species.

The following abbreviations are used herein: bp(s), base pair(s); DEAE, *O*-diethylaminoethyl; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; GC, gas chromatography; IPTG, isopropyl-β-D-thiogalactopyranoside; LB, Luria-Bertani; Mopso, 3-(*N*-morpholino)-2-hydroxypropane-sulfonic acid; MS, mass spectrum/ spectrometry; nt(s), nucleotide(s); ORF, open reading frame; PCR,

polymerase chain reaction; PVDF, polyvinylidenedifluoride; SDS, sodium dodecyl sulfate; Tris, tris(hydroxymethyl) aminomethane; UV, ultraviolet.

In accordance with the present invention, cDNAs encoding myrcene synthase (SEQ ID NO:1), (-)-pinene synthase (SEQ ID NO:3) and (-)-limonene synthase (SEQ ID NO:5) from Grand fir (*Abies grandis*) were isolated and sequenced in the following manner. Based on comparison of sequences of limonene synthase from spearmint (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024), 5-*epi*-aristolochene synthase from tobacco (Facchini, P.J., and Chappell, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**:11088-11092), and casbene synthase from castor bean (Mau, C.J.D., and West, C.A. (1994) *Proc. Natl. Acad. Sci. USA* **91**:8497-8501), four conserved regions were identified for which a set of consensus, degenerate primers were synthesized: Primer A (SEQ ID NO:7), Primer B (SEQ ID NO:8), Primer C (SEQ ID NO:9) and Primer D (SEQ ID NO:10). Primers A (SEQ ID NO:7), B (SEQ ID NO:8), and D (SEQ ID NO:10) were sense primers, while Primer C (SEQ ID NO:9), was an antisense primer. Each of the sense primers, A (SEQ ID NO:7), B (SEQ ID NO:8) and D (SEQ ID NO:10), was used for PCR in combination with antisense primer C (SEQ ID NO:9) by employing a broad range of amplification conditions. Analysis of the PCR reaction products by agarose gel electrophoresis revealed that only the combination of primers C (SEQ ID NO:9) and D (SEQ ID NO:10) generated a specific PCR product of approximately 110 bps.

The 110 bps PCR product was gel purified, ligated into a plasmid, and transformed into *E. coli* XL1-Blue cells. Plasmid DNA was prepared from 41 individual transformants and the inserts were sequenced. Four different insert sequences were identified, and were designated as probes 1 (SEQ ID NO:11), 2 (SEQ ID NO:12), 4 (SEQ ID NO:13) and 5 (SEQ ID NO:14). Probes 1 (SEQ ID NO:11), 2 (SEQ ID NO:12), 4 (SEQ ID NO:13) and 5 (SEQ ID NO:14) were used to screen a cDNA library made from mRNA extracted from wounded Grand fir stems, and the longest clone that hybridized to each of these probes was isolated and sequenced. Thus, clone *AG1.28* (SEQ ID NO:15) is the longest cDNA clone that hybridized to probe 1 (SEQ ID NO:11), clone *AG2.2* (SEQ ID NO:1) is the longest cDNA clone that hybridized to probe 2 (SEQ ID NO:12), clone *AG4.30* (SEQ ID NO:17) is the longest cDNA clone that hybridized to probe 4 (SEQ ID NO: 13), and clone *AG5.9* (SEQ ID NO:19) is the longest cDNA clone that hybridized to probe 5 (SEQ ID NO:14).

Truncated clone *AG1.28* (SEQ ID NO:15) resembled most closely in size and sequence (72% similarity, 49% identity) a diterpene cyclase, abietadiene synthase, from Grand fir. Clones *AG4.30* (SEQ ID NO:17) and *AG5.9* (SEQ ID NO:19) encode sesquiterpene synthases. Sequence and functional analysis of clone *AG2.2* (SEQ ID NO:1) revealed that it encoded the monoterpene synthase, myrcene synthase.

Alignment of the four new terpene synthase cDNA sequences *AG1.28* (SEQ ID NO:15), *AG2.2* (SEQ ID NO:1), *AG4.30* (SEQ ID NO:17) and *AG5.9* (SEQ ID NO:19) with that for abietadiene synthase (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268) allowed the identification of several, conserved sequence motifs. Two new sense PCR primers, primer E (SEQ ID NO:21) and primer F (SEQ ID NO:22) were designed based on the sequence of the conserved protein sequence motifs. A new antisense PCR primer, primer G (SEQ ID NO:23), was designed based on limited sequence information available from pinene synthase. The combination of primer E (SEQ ID NO:21) and primer G (SEQ ID NO:23) amplified a cDNA product of 1022 bps, which was designated as probe 3 (SEQ ID NO:24).

Probe 3 (SEQ ID NO:24) was used to screen a cDNA library made from mRNA extracted from wounded Grand fir stems. Hybridization of 10^5 Grand fir λZAP II cDNA clones with probe 3 (SEQ ID NO:24) yielded two types of signals comprised of about 400 strongly positive clones and an equal number of weak positives, indicating that the probe recognized more than one type of cDNA. Thirty-four of the former clones and eighteen of the latter were purified, the inserts were selected by size (2.0-2.5 kb), and the *in vivo* excised clones were partially sequenced from both ends. Those clones which afforded weak hybridization signals were shown to contain inserts that were either identical to myrcene synthase clone *AG2.2* (SEQ ID NO:1) or exhibited no significant sequence similarity to terpene synthases.

Clones which gave strong hybridization signals segregated into distinct sequence groups represented by clone *AG3.18* (SEQ ID NO:3) and clone *AG10* (SEQ ID NO:5). Both *AG3.18* (SEQ ID NO:3) and *AG10* (SEQ ID NO:5) were subcloned into plasmid expression vectors and expressed in *E. coli*. When extracts of the induced cells were tested for terpene synthase activity with all of the potential prenyl diphosphate substrates, only geranyl diphosphate was utilized. Extracts from *E. coli* containing the *AG10* (SEQ ID NO:5) expression construct converted geranyl diphosphate to the -4S enantiomer of limonene as the major product, indicating that

AG10 (SEQ ID NO:5) encodes (-)-limonene synthase. Similar analysis of the monoterpene products generated from geranyl diphosphate by cell-free extracts of *E. coli* containing the *AG3.18* (SEQ ID NO:3) insert ligated into an expression vector revealed the presence of a 42:58% mixture of α-pinene and β-pinene, the same product ratio previously described for the purified, native (-)-pinene synthase from Grand fir. These data indicate that *AG3.18* (SEQ ID NO:3) encodes (-)-pinene synthase.

The isolation of cDNAs encoding (-)-limonene synthase, (-)-pinene synthase and myrcene synthase permits the development of efficient expression systems for these functional enzymes; provides useful tools for examining the developmental regulation of monoterpene biosynthesis; permits investigation of the reaction mechanism(s) of these unusual, multiproduct enzymes, and permits the isolation of other (-)-limonene synthase, (-)-pinene synthase and myrcene synthases. The isolation of the (-)-limonene synthase, (-)-pinene synthase and myrcene synthase cDNAs also permits the transformation of a wide range of organisms in order to introduce monoterpene biosynthesis *de novo*, or to modify endogenous monoterpene biosynthesis.

Although the (-)-limonene synthase, (-)-pinene synthase and myrcene synthase proteins set forth in SEQ ID NO:6, SEQ ID NO:4 and SEQ ID NO:2, respectively, direct the enzymes to plastids, substitution of the presumptive targeting sequence of each of these enzymes (SEQ ID NO:2, amino acids 1 to 61; SEQ ID NO:4, amino acids 1 to 61; SEQ ID NO:6, amino acids 1 to 66) with other transport sequences well known in the art (see, e.g., von Heijne et al., *Eur. J. Biochem.* **180**:535-545, 1989; Stryer, *Biochemistry*, W.H. Freeman and Company, New York, NY, p. 769 [1988]) may be employed to direct the (-)-limonene synthase, (-)-pinene synthase and myrcene synthase to other cellular or extracellular locations.

In addition to the native (-)-limonene synthase, (-)-pinene synthase and myrcene synthase amino acid sequences of SEQ ID NO:6, SEQ ID NO:4 and SEQ ID NO:2, respectively, sequence variants produced by deletions, substitutions, mutations and/or insertions are intended to be within the scope of the invention except insofar as limited by the prior art. The (-)-limonene synthase, (-)-pinene synthase and myrcene synthase amino acid sequence variants of this invention may be constructed by mutating the DNA sequences that encode the wild-type synthases, such as by using techniques commonly referred to as site-directed mutagenesis. Nucleic acid molecules encoding the monoterpene synthases of the present invention

can be mutated by a variety of PCR techniques well known to one of ordinary skill in the art. *See, e.g.*, "PCR Strategies", M.A. Innis, D.H. Gelfand and J.J. Sninsky, eds., 1995, Academic Press, San Diego, CA (Chapter 14); "PCR Protocols: A Guide to Methods and Applications", M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White, eds., Academic Press, NY (1990).

By way of non-limiting example, the two primer system utilized in the Transformer Site-Directed Mutagenesis kit from Clontech, may be employed for introducing site-directed mutants into the monoterpene synthase genes of the present invention. Following denaturation of the target plasmid in this system, two primers are simultaneously annealed to the plasmid; one of these primers contains the desired site-directed mutation, the other contains a mutation at another point in the plasmid resulting in elimination of a restriction site. Second strand synthesis is then carried out, tightly linking these two mutations, and the resulting plasmids are transformed into a *mutS* strain of *E. coli*. Plasmid DNA is isolated from the transformed bacteria, restricted with the relevant restriction enzyme (thereby linearizing the unmutated plasmids), and then retransformed into *E. coli*. This system allows for generation of mutations directly in an expression plasmid, without the necessity of subcloning or generation of single-stranded phagemids. The tight linkage of the two mutations and the subsequent linearization of unmutated plasmids results in high mutation efficiency and allows minimal screening. Following synthesis of the initial restriction site primer, this method requires the use of only one new primer type per mutation site. Rather than prepare each positional mutant separately, a set of "designed degenerate" oligonucleotide primers can be synthesized in order to introduce all of the desired mutations at a given site simultaneously. Transformants can be screened by sequencing the plasmid DNA through the mutagenized region to identify and sort mutant clones. Each mutant DNA can then be restricted and analyzed by electrophoresis on Mutation Detection Enhancement gel (J.T. Baker) to confirm that no other alterations in the sequence have occurred (by band shift comparison to the unmutagenized control).

The verified mutant duplexes in the pET (or other) overexpression vector can be employed to transform *E. coli* such as strain *E. coli* BL21(DE3)pLysS, for high level production of the mutant protein, and purification by standard protocols. The method of FAB-MS mapping can be employed to rapidly check the fidelity of mutant expression. This technique provides for sequencing segments throughout the whole protein and provides the necessary confidence in the sequence assignment. In a

mapping experiment of this type, protein is digested with a protease (the choice will depend on the specific region to be modified since this segment is of prime interest and the remaining map should be identical to the map of unmutagenized protein). The set of cleavage fragments is fractionated by microbore HPLC (reversed phase or ion exchange, again depending on the specific region to be modified) to provide several peptides in each fraction, and the molecular weights of the peptides are determined by FAB-MS. The masses are then compared to the molecular weights of peptides expected from the digestion of the predicted sequence, and the correctness of the sequence quickly ascertained. Since this mutagenesis approach to protein modification is directed, sequencing of the altered peptide should not be necessary if the MS agrees with prediction. If necessary to verify a changed residue, CAD-tandem MS/MS can be employed to sequence the peptides of the mixture in question, or the target peptide purified for subtractive Edman degradation or carboxypeptidase Y digestion depending on the location of the modification.

In the design of a particular site directed mutagenesis, it is generally desirable to first make a non-conservative substitution (e.g., Ala for Cys, His or Glu) and determine if activity is greatly impaired as a consequence. The properties of the mutagenized protein are then examined with particular attention to the kinetic parameters of K_m and k_{cat} as sensitive indicators of altered function, from which changes in binding and/or catalysis *per se* may be deduced by comparison to the native enzyme. If the residue is by this means demonstrated to be important by activity impairment, or knockout, then conservative substitutions can be made, such as Asp for Glu to alter side chain length, Ser for Cys, or Arg for His. For hydrophobic segments, it is largely size that is usefully altered, although aromatics can also be substituted for alkyl side chains. Changes in the normal product distribution can indicate which step(s) of the reaction sequence have been altered by the mutation. Modification of the hydrophobic pocket can be employed to change binding conformations for substrates and result in altered regiochemistry and/or stereochemistry.

Other site directed mutagenesis techniques may also be employed with the nucleotide sequences of the invention. For example, restriction endonuclease digestion of DNA followed by ligation may be used to generate deletion variants of (-)-limonene synthase, (-)-pinene synthase and myrcene synthase, as described in section 15.3 of Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, New York, NY [1989]). A similar strategy

may be used to construct insertion variants, as described in section 15.3 of Sambrook et al., *supra*.

Oligonucleotide-directed mutagenesis may also be employed for preparing substitution variants of this invention. It may also be used to conveniently prepare the deletion and insertion variants of this invention. This technique is well known in the art as described by Adelman et al. (*DNA* 2:183 [1983]); Sambrook et al., *supra*; "Current Protocols in Molecular Biology", 1991, Wiley (NY), F.T. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.D. Seidman, J.A. Smith and K. Struhl, eds.

Generally, oligonucleotides of at least 25 nucleotides in length are used to insert, delete or substitute two or more nucleotides in the (-)-limonene synthase, (-)-pinene synthase and myrcene synthase molecule. An optimal oligonucleotide will have 12 to 15 perfectly matched nucleotides on either side of the nucleotides coding for the mutation. To mutagenize wild-type (-)-limonene synthase, (-)-pinene synthase and myrcene synthase, the oligonucleotide is annealed to the single-stranded DNA template molecule under suitable hybridization conditions. A DNA polymerizing enzyme, usually the Klenow fragment of *E. coli* DNA polymerase I, is then added. This enzyme uses the oligonucleotide as a primer to complete the synthesis of the mutation-bearing strand of DNA. Thus, a heteroduplex molecule is formed such that one strand of DNA encodes the wild-type synthase inserted in the vector, and the second strand of DNA encodes the mutated form of the synthase inserted into the same vector. This heteroduplex molecule is then transformed into a suitable host cell.

Mutants with more than one amino acid substituted may be generated in one of several ways. If the amino acids are located close together in the polypeptide chain, they may be mutated simultaneously using one oligonucleotide that codes for all of the desired amino acid substitutions. If however, the amino acids are located some distance from each other (separated by more than ten amino acids, for example) it is more difficult to generate a single oligonucleotide that encodes all of the desired changes. Instead, one of two alternative methods may be employed. In the first method, a separate oligonucleotide is generated for each amino acid to be substituted. The oligonucleotides are then annealed to the single-stranded template DNA simultaneously, and the second strand of DNA that is synthesized from the template will encode all of the desired amino acid substitutions. An alternative method involves two or more rounds of mutagenesis to produce the desired mutant. The first round is as described for the single mutants: wild-type (-)-limonene synthase,

(-)-pinene synthase and myrcene synthase DNA is used for the template, an oligonucleotide encoding the first desired amino acid substitution(s) is annealed to this template, and the heteroduplex DNA molecule is then generated. The second round of mutagenesis utilizes the mutated DNA produced in the first round of mutagenesis as the template. Thus, this template already contains one or more mutations. The oligonucleotide encoding the additional desired amino acid substitution(s) is then annealed to this template, and the resulting strand of DNA now encodes mutations from both the first and second rounds of mutagenesis. This resultant DNA can be used as a template in a third round of mutagenesis, and so on.

A gene encoding (-)-limonene synthase, (-)-pinene synthase and myrcene synthase may be incorporated into any organism (intact plant, animal, microbe, etc.), or cell culture derived therefrom, that produces geranyl diphosphate. A (-)-limonene synthase, (-)-pinene synthase and myrcene synthase gene may be introduced into any organism for a variety of purposes including, but not limited to: production of (-)-limonene synthase, (-)-pinene synthase and myrcene synthase, or their products; production or modification of flavor and aroma properties; improvement of defense capability, and the alteration of other ecological interactions mediated by myrcene, (-)-limonene, (-)-pinene, or their derivatives.

Eukaryotic expression systems may be utilized for the production of (-)-limonene synthase, (-)-pinene synthase and myrcene synthase since they are capable of carrying out any required posttranslational modifications and of directing the enzymes to the proper membrane location. A representative eukaryotic expression system for this purpose uses the recombinant baculovirus, *Autographa californica* nuclear polyhedrosis virus (AcNPV; M.D. Summers and G.E. Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures* [1986]; Luckow et al., *Bio-technology* **6**:47-55 [1987]) for expression of the terpenoid synthases of the invention. Infection of insect cells (such as cells of the species *Spodoptera frugiperda*) with the recombinant baculoviruses allows for the production of large amounts of the monoterpenoid synthase proteins. In addition, the baculovirus system has other important advantages for the production of recombinant (-)-limonene synthase, (-)-pinene synthase and myrcene synthase. For example, baculoviruses do not infect humans and can therefore be safely handled in large quantities. In the baculovirus system, a DNA construct is prepared including a DNA segment encoding (-)-limonene synthase, (-)-pinene synthase and myrcene synthase and a vector. The vector may comprise the polyhedron gene promoter region of a

baculovirus, the baculovirus flanking sequences necessary for proper cross-over during recombination (the flanking sequences comprise about 200-300 base pairs adjacent to the promoter sequence) and a bacterial origin of replication which permits the construct to replicate in bacteria. The vector is constructed so that (i) the DNA segment is placed adjacent (or operably linked or "downstream" or "under the control of") to the polyhedron gene promoter and (ii) the promoter/monoterpene synthase combination is flanked on both sides by 200-300 base pairs of baculovirus DNA (the flanking sequences).

To produce the monoterpene synthase DNA construct, a cDNA clone encoding the full length (-)-limonene synthase, (-)-pinene synthase and myrcene synthase is obtained using methods such as those described herein. The DNA construct is contacted in a host cell with baculovirus DNA of an appropriate baculovirus (that is, of the same species of baculovirus as the promoter encoded in the construct) under conditions such that recombination is effected. The resulting recombinant baculoviruses encode the full (-)-limonene synthase, (-)-pinene synthase and myrcene synthase. For example, an insect host cell can be cotransfected or transfected separately with the DNA construct and a functional baculovirus. Resulting recombinant baculoviruses can then be isolated and used to infect cells to effect production of the monoterpene synthase. Host insect cells include, for example, *Spodoptera frugiperda* cells, that are capable of producing a baculovirus-expressed monoterpene synthase. Insect host cells infected with a recombinant baculovirus of the present invention are then cultured under conditions allowing expression of the baculovirus-encoded (-)-limonene synthase, (-)-pinene synthase and myrcene synthase. (-)-limonene synthase, (-)-pinene synthase and myrcene synthase thus produced are then extracted from the cells using methods known in the art.

Other eukaryotic microbes such as yeasts may also be used to practice this invention. The baker's yeast *Saccharomyces cerevisiae*, is a commonly used yeast, although several other strains are available. The plasmid YRp7 (Stinchcomb et al., *Nature* **282**:39 [1979]; Kingsman et al., *Gene* 7:141 [1979]; Tschemper et al., *Gene* **10**:157 [1980]) is commonly used as an expression vector in *Saccharomyces*. This plasmid contains the trp1 gene that provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, such as strains ATCC No. 44,076 and PEP4-1 (Jones, *Genetics* **85**:12 [1977]). The presence of the trp1 lesion as a characteristic of the yeast host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan. Yeast host cells

are generally transformed using the polyethylene glycol method, as described by Hinnen (*Proc. Natl. Acad. Sci. USA* **75**:1929 [1978]). Additional yeast transformation protocols are set forth in Gietz et al., *N.A.R.* **20(17)** 1425(1992); Reeves et al., *FEMS* **99(2-3)**:193-197, (1992).

Suitable promoting sequences in yeast vectors include the promoters for 3-phosphoglycerate kinase (Hitzeman et al., *J. Biol. Chem.* **255**:2073 [1980]) or other glycolytic enzymes (Hess et al., *J. Adv. Enzyme Reg.* **7**:149 [1968]; Holland et al., *Biochemistry* **17**:4900 [1978]), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. In the construction of suitable expression plasmids, the termination sequences associated with these genes are also ligated into the expression vector 3' of the sequence desired to be expressed to provide polyadenylation of the mRNA and termination. Other promoters that have the additional advantage of transcription controlled by growth conditions are the promoter region for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Any plasmid vector containing yeast-compatible promoter, origin of replication and termination sequences is suitable.

Cell cultures derived from multicellular organisms, such as plants, may be used as hosts to practice this invention. Transgenic plants can be obtained, for example, by transferring plasmids that encode (-)-limonene synthase, (-)-pinene synthase and myrcene synthase and a selectable marker gene, e.g., the kan gene encoding resistance to kanamycin, into *Agrobacterium tumifaciens* containing a helper Ti plasmid as described in Hoeckema et al., *Nature* **303**:179-181 [1983] and culturing the *Agrobacterium* cells with leaf slices of the plant to be transformed as described by An et al., *Plant Physiology* **81**:301-305 [1986]. Transformation of cultured plant host cells is normally accomplished through *Agrobacterium tumifaciens*, as described above. Cultures of mammalian host cells and other host cells that do not have rigid cell membrane barriers are usually transformed using the calcium phosphate method as originally described by Graham and Van der Eb (*Virology* **52**:546 [1978]) and modified as described in sections 16.32-16.37 of Sambrook et al., *supra*. However, other methods for introducing DNA into cells

such as Polybrene (Kawai and Nishizawa, *Mol. Cell. Biol.* 4:1172 [1984]), protoplast fusion (Schaffner, *Proc. Natl. Acad. Sci. USA* 77:2163 [1980]), electroporation (Neumann et al., *EMBO J.* 1:841 [1982]), and direct microinjection into nuclei (Capecchi, *Cell* 22:479 [1980]) may also be used. Additionally, animal transformation strategies are reviewed in Monastersky G.M. and Robl, J.M., *Strategies in Transgenic Animal Science*, ASM Press, Washington, D.C., 1995. Transformed plant calli may be selected through the selectable marker by growing the cells on a medium containing, e.g., kanamycin, and appropriate amounts of phytohormone such as naphthalene acetic acid and benzyladenine for callus and shoot induction. The plant cells may then be regenerated and the resulting plants transferred to soil using techniques well known to those skilled in the art.

In addition, a gene regulating (-)-limonene synthase, (-)-pinene synthase and myrcene synthase production can be incorporated into the plant along with a necessary promoter which is inducible. In the practice of this embodiment of the invention, a promoter that only responds to a specific external or internal stimulus is fused to the target cDNA. Thus, the gene will not be transcribed except in response to the specific stimulus. As long as the gene is not being transcribed, its gene product is not produced.

An illustrative example of a responsive promoter system that can be used in the practice of this invention is the glutathione-S-transferase (GST) system in maize. GSTs are a family of enzymes that can detoxify a number of hydrophobic electrophilic compounds that often are used as pre-emergent herbicides (Weigand et al., *Plant Molecular Biology* 7:235-243 [1986]). Studies have shown that the GSTs are directly involved in causing this enhanced herbicide tolerance. This action is primarily mediated through a specific 1.1 kb mRNA transcription product. In short, maize has a naturally occurring quiescent gene already present that can respond to external stimuli and that can be induced to produce a gene product. This gene has previously been identified and cloned. Thus, in one embodiment of this invention, the promoter is removed from the GST responsive gene and attached to a (-)-limonene synthase, (-)-pinene synthase and myrcene synthase gene that previously has had its native promoter removed. This engineered gene is the combination of a promoter that responds to an external chemical stimulus and a gene responsible for successful production of (-)-limonene synthase, (-)-pinene synthase and myrcene synthase.

In addition to the methods described above, several methods are known in the art for transferring cloned DNA into a wide variety of plant species, including gymnosperms, angiosperms, monocots and dicots (see, e.g., Glick and Thompson, eds., *Methods in Plant Molecular Biology*, CRC Press, Boca Raton, Florida [1993]). Representative examples include electroporation-facilitated DNA uptake by protoplasts (Rhodes et al., *Science* **240(4849):**204-207 [1988]); treatment of protoplasts with polyethylene glycol (Lyznik et al., *Plant Molecular Biology* **13:**151-161 [1989]); and bombardment of cells with DNA laden microprojectiles (Klein et al., *Plant Physiol.* **91:**440-444 [1989] and Boynton et al., *Science* **240(4858):**1534-1538 [1988]). Additionally, plant transformation strategies and techniques are reviewed in Birch, R.G., *Ann Rev Plant Phys Plant Mol Biol* **48:**297 (1997); Forester et al., *Exp. Agric.* **33:**15-33 (1997). Minor variations make these technologies applicable to a broad range of plant species.

Each of these techniques has advantages and disadvantages. In each of the techniques, DNA from a plasmid is genetically engineered such that it contains not only the gene of interest, but also selectable and screenable marker genes. A selectable marker gene is used to select only those cells that have integrated copies of the plasmid (the construction is such that the gene of interest and the selectable and screenable genes are transferred as a unit). The screenable gene provides another check for the successful culturing of only those cells carrying the genes of interest. A commonly used selectable marker gene is neomycin phosphotransferase II (NPT II). This gene conveys resistance to kanamycin, a compound that can be added directly to the growth media on which the cells grow. Plant cells are normally susceptible to kanamycin and, as a result, die. The presence of the NPT II gene overcomes the effects of the kanamycin and each cell with this gene remains viable. Another selectable marker gene which can be employed in the practice of this invention is the gene which confers resistance to the herbicide glufosinate (Basta). A screenable gene commonly used is the β-glucuronidase gene (GUS). The presence of this gene is characterized using a histochemical reaction in which a sample of putatively transformed cells is treated with a GUS assay solution. After an appropriate incubation, the cells containing the GUS gene turn blue.

The plasmid containing one or more of these genes is introduced into either plant protoplasts or callus cells by any of the previously mentioned techniques. If the marker gene is a selectable gene, only those cells that have incorporated the DNA package survive under selection with the appropriate phytotoxic agent. Once the

appropriate cells are identified and propagated, plants are regenerated. Progeny from the transformed plants must be tested to insure that the DNA package has been successfully integrated into the plant genome.

Mammalian host cells may also be used in the practice of the invention. Examples of suitable mammalian cell lines include monkey kidney CVI line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line 293S (Graham et al., *J. Gen. Virol.* **36**:59 [1977]); baby hamster kidney cells (BHK, ATCC CCL 10); Chinese hamster ovary cells (Urlab and Chasin, *Proc. Natl. Acad. Sci USA* **77**:4216 [1980]); mouse sertoli cells (TM4, Mather, *Biol. Reprod.* **23**:243 [1980]); monkey kidney cells (CVI-76, ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1587); human cervical carcinoma cells (HELA, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor cells (MMT 060562, ATCC CCL 51); rat hepatoma cells (HTC, MI.54, Baumann et al., *J. Cell Biol.* **85**:1 [1980]); and TRI cells (Mather et al., *Annals N.Y. Acad. Sci.* **383**:44 [1982]). Expression vectors for these cells ordinarily include (if necessary) DNA sequences for an origin of replication, a promoter located in front of the gene to be expressed, a ribosome binding site, an RNA splice site, a polyadenylation site, and a transcription terminator site.

Promoters used in mammalian expression vectors are often of viral origin. These viral promoters are commonly derived from polyoma virus, Adenovirus 2, and most frequently Simian Virus 40 (SV40). The SV40 virus contains two promoters that are termed the early and late promoters. These promoters are particularly useful because they are both easily obtained from the virus as one DNA fragment that also contains the viral origin of replication (Fiers et al., *Nature* **273**:113 [1978]). Smaller or larger SV40 DNA fragments may also be used, provided they contain the approximately 250-bp sequence extending from the HindIII site toward the BglI site located in the viral origin of replication.

Alternatively, promoters that are naturally associated with the foreign gene (homologous promoters) may be used provided that they are compatible with the host cell line selected for transformation.

An origin of replication may be obtained from an exogenous source, such as SV40 or other virus (e.g., Polyoma, Adeno, VSV, BPV) and inserted into the cloning vector. Alternatively, the origin of replication may be provided by the host cell

chromosomal replication mechanism. If the vector containing the foreign gene is integrated into the host cell chromosome, the latter is often sufficient.

The use of a secondary DNA coding sequence can enhance production levels of (-)-limonene synthase, (-)-pinene synthase and myrcene synthase in transformed cell lines. The secondary coding sequence typically comprises the enzyme dihydrofolate reductase (DHFR). The wild-type form of DHFR is normally inhibited by the chemical methotrexate (MTX). The level of DHFR expression in a cell will vary depending on the amount of MTX added to the cultured host cells. An additional feature of DHFR that makes it particularly useful as a secondary sequence is that it can be used as a selection marker to identify transformed cells. Two forms of DHFR are available for use as secondary sequences, wild-type DHFR and MTX-resistant DHFR. The type of DHFR used in a particular host cell depends on whether the host cell is DHFR deficient (such that it either produces very low levels of DHFR endogenously, or it does not produce functional DHFR at all). DHFR-deficient cell lines such as the CHO cell line described by Urlaub and Chasin, *supra*, are transformed with wild-type DHFR coding sequences. After transformation, these DHFR-deficient cell lines express functional DHFR and are capable of growing in a culture medium lacking the nutrients hypoxanthine, glycine and thymidine. Nontransformed cells will not survive in this medium.

The MTX-resistant form of DHFR can be used as a means of selecting for transformed host cells in those host cells that endogenously produce normal amounts of functional DHFR that is MTX sensitive. The CHO-K1 cell line (ATCC No. CL 61) possesses these characteristics, and is thus a useful cell line for this purpose. The addition of MTX to the cell culture medium will permit only those cells transformed with the DNA encoding the MTX-resistant DHFR to grow. The nontransformed cells will be unable to survive in this medium.

Prokaryotes may also be used as host cells for the initial cloning steps of this invention. They are particularly useful for rapid production of large amounts of DNA, for production of single-stranded DNA templates used for site-directed mutagenesis, for screening many mutants simultaneously, and for DNA sequencing of the mutants generated. Suitable prokaryotic host cells include *E. coli* K12 strain 94 (ATCC No. 31,446), *E. coli* strain W3110 (ATCC No. 27,325) *E. coli* X1776 (ATCC No. 31,537), and *E. coli* B; however many other strains of *E. coli*, such as HB101, JM101, NM522, NM538, NM539, and many other species and genera of prokaryotes including bacilli such as *Bacillus subtilis*, other

enterobacteriaceae such as *Salmonella typhimurium* or *Serratia marcesans*, and various *Pseudomonas* species may all be used as hosts. Prokaryotic host cells or other host cells with rigid cell walls are preferably transformed using the calcium chloride method as described in section 1.82 of Sambrook et al., *supra*. Alternatively, electroporation may be used for transformation of these cells. Prokaryote transformation techniques are set forth in Dower, W.J., in Genetic Engineering, Principles and Methods, **12**:275-296, Plenum Publishing Corp., 1990; Hanahan et al., *Meth. Enxymol.*, **204**:63 (1991).

As a representative example, cDNA sequences encoding (-)-limonene synthase, (-)-pinene synthase or myrcene synthase may be transferred to the $(His)_6$•Tag pET vector commercially available (from Novagen) for overexpression in *E. coli* as heterologous host. This pET expression plasmid has several advantages in high level heterologous expression systems. The desired cDNA insert is ligated in frame to plasmid vector sequences encoding six histidines followed by a highly specific protease recognition site (thrombin) that are joined to the amino terminus codon of the target protein. The histidine "block" of the expressed fusion protein promotes very tight binding to immobilized metal ions and permits rapid purification of the recombinant protein by immobilized metal ion affinity chromatography. The histidine leader sequence is then cleaved at the specific proteolysis site by treatment of the purified protein with thrombin, and the (-)-limonene synthase, (-)-pinene synthase and myrcene synthase again purified by immobilized metal ion affinity chromatography, this time using a shallower imidazole gradient to elute the recombinant synthases while leaving the histidine block still adsorbed. This overexpression-purification system has high capacity, excellent resolving power and is fast, and the chance of a contaminating *E. coli* protein exhibiting similar binding behavior (before and after thrombin proteolysis) is extremely small.

As will be apparent to those skilled in the art, any plasmid vectors containing replicon and control sequences that are derived from species compatible with the host cell may also be used in the practice of the invention. The vector usually has a replication site, marker genes that provide phenotypic selection in transformed cells, one or more promoters, and a polylinker region containing several restriction sites for insertion of foreign DNA. Plasmids typically used for transformation of *E. coli* include pBR322, pUC18, pUC19, pUCI18, pUC119, and Bluescript M13, all of which are described in sections 1.12-1.20 of Sambrook et al., *supra*. However, many other suitable vectors are available as well. These vectors contain genes coding for

ampicillin and/or tetracycline resistance which enables cells transformed with these vectors to grow in the presence of these antibiotics.

The promoters most commonly used in prokaryotic vectors include the β-lactamase (penicillinase) and lactose promoter systems (Chang et al. *Nature* 375:615 [1978]; Itakura et al., *Science* **198**:1056 [1977]; Goeddel et al., *Nature* **281**:544 [1979]) and a tryptophan (trp) promoter system (Goeddel et al., *Nucl. Acids Res.* **8**:4057 [1980]; EPO Appl. Publ. No. 36,776), and the alkaline phosphatase systems. While these are the most commonly used, other microbial promoters have been utilized, and details concerning their nucleotide sequences have been published, enabling a skilled worker to ligate them functionally into plasmid vectors (see Siebenlist et al., *Cell* **20**:269 [1980]).

Many eukaryotic proteins normally secreted from the cell contain an endogenous secretion signal sequence as part of the amino acid sequence. Thus, proteins normally found in the cytoplasm can be targeted for secretion by linking a signal sequence to the protein. This is readily accomplished by ligating DNA encoding a signal sequence to the 5' end of the DNA encoding the protein and then expressing this fusion protein in an appropriate host cell. The DNA encoding the signal sequence may be obtained as a restriction fragment from any gene encoding a protein with a signal sequence. Thus, prokaryotic, yeast, and eukaryotic signal sequences may be used herein, depending on the type of host cell utilized to practice the invention. The DNA and amino acid sequence encoding the signal sequence portion of several eukaryotic genes including, for example, human growth hormone, proinsulin, and proalbumin are known (see Stryer, *Biochemistry* W.H. Freeman and Company, New York, NY, p. 769 [1988]), and can be used as signal sequences in appropriate eukaryotic host cells. Yeast signal sequences, as for example acid phosphatase (Arima et al., *Nuc. Acids Res.* **11**:1657 [1983]), α–factor, alkaline phosphatase and invertase may be used to direct secretion from yeast host cells. Prokaryotic signal sequences from genes encoding, for example, LamB or OmpF (Wong et al., *Gene* **68**:193 [1988]), MalE, PhoA, or beta-lactamase, as well as other genes, may be used to target proteins from prokaryotic cells into the culture medium.

Trafficking sequences from plants, animals and microbes can be employed in the practice of the invention to direct the monoterpene synthase proteins of the present invention to the cytoplasm, endoplasmic reticulum, mitochondria or other cellular components, or to target the protein for export to the medium. These considerations apply to the overexpression of (-)-limonene synthase, (-)-pinene

synthase and myrcene synthase, and to direction of expression within cells or intact organisms to permit gene product function in any desired location.

The construction of suitable vectors containing DNA encoding replication sequences, regulatory sequences, phenotypic selection genes and the monoterpene synthase DNA of interest are prepared using standard recombinant DNA procedures. Isolated plasmids and DNA fragments are cleaved, tailored, and ligated together in a specific order to generate the desired vectors, as is well known in the art (see, for example, Maniatis, *supra*, and Sambrook et al., *supra*).

As discussed above, (-)-limonene synthase, (-)-pinene synthase and myrcene synthase variants are preferably produced by means of mutation(s) that are generated using the method of site-specific mutagenesis. This method requires the synthesis and use of specific oligonucleotides that encode both the sequence of the desired mutation and a sufficient number of adjacent nucleotides to allow the oligonucleotide to stably hybridize to the DNA template.

The foregoing may be more fully understood in connection with the following representative examples, in which "Plasmids" are designated by a lower case p followed by an alphanumeric designation. The starting plasmids used in this invention are either commercially available, publicly available on an unrestricted basis, or can be constructed from such available plasmids using published procedures. In addition, other equivalent plasmids are known in the art and will be apparent to the ordinary artisan.

"Digestion", "cutting" or "cleaving" of DNA refers to catalytic cleavage of the DNA with an enzyme that acts only at particular locations in the DNA. These enzymes are called restriction endonucleases, and the site along the DNA sequence where each enzyme cleaves is called a restriction site. The restriction enzymes used in this invention are commercially available and are used according to the instructions supplied by the manufacturers. (See also sections 1.60-1.61 and sections 3.38-3.39 of Sambrook et al., *supra*.)

"Recovery" or "isolation" of a given fragment of DNA from a restriction digest means separation of the resulting DNA fragment on a polyacrylamide or an agarose gel by electrophoresis, identification of the fragment of interest by comparison of its mobility versus that of marker DNA fragments of known molecular weight, removal of the gel section containing the desired fragment, and separation of the gel from DNA. This procedure is known generally. For example, see Lawn et al.

(*Nucleic Acids Res.* **9**:6103-6114 [1982]), and Goeddel et al. (*Nucleic Acids Res., supra*).

The following examples merely illustrate the best mode now contemplated for practicing the invention, but should not be construed to limit the invention. All literature citations herein are expressly incorporated by reference.

EXAMPLE 1

PCR-Based Generation of Probes for Cloning Monoterpene Synthases From Grand fir (*Abies grandis*)

Substrates, Reagents and cDNA Library – [1-^{3}H]Geranyl diphosphate (250 Ci/mol) (Croteau, R., Alonso, W.R., Koepp, A.E., and Johnson, M.A. (1994) *Arch. Biochem. Biophys.* **309**:184-192), [1-^{3}H]farnesyl diphosphate (125 Ci/mol) (Dehal, S.S., and Croteau, R. (1988) *Arch. Biochem. Biophys.* **261**:346-356) and [1-^{3}H]geranylgeranyl diphosphate (120 Ci/mol) (LaFever, R.E., Stofer Vogel, B., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **313**:139-149) were prepared as described previously. Terpenoid standards were from our own collection. All other biochemicals and reagents were purchased from Sigma Chemical Co. or Aldrich Chemical Co., unless otherwise noted. Construction of the λZAP II cDNA library, using mRNA isolated from wounded Grand fir sapling stems, was described previously (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268).

PCR-Based Probe Generation – Based on comparison of sequences of limonene synthase from spearmint (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024), 5-*epi*-aristolochene synthase from tobacco (Facchini, P.J., and Chappell, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**:11088-11092), and casbene synthase from castor bean (Mau, C.J.D., and West, C.A. (1994) *Proc. Natl. Acad. Sci. USA* **91**:8497-8501), four conserved regions were identified for which a set of consensus degenerate primers were synthesized: Primer A (SEQ ID NO:7); Primer B (SEQ ID NO:8); Primer C (SEQ ID NO:9); Primer D (SEQ ID NO:10). Primers A (SEQ ID NO:7), B (SEQ ID NO:8) and C (SEQ ID NO:9) have been described previously (Steele, C., Lewinsohn, E. and Croteau, R., *Proc. Nat'l. Acad. Sci. USA*, **92:** 4164-4168 (1995)); primer D (SEQ ID NO:10) was designed based on the conserved amino acid sequence motif DD(T/I)(I/Y/F)D(A/V)Y(A/G)(SEQ ID NO:25) of the above noted terpene synthases (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024;

Facchini, P.J., and Chappell, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**:11088-11092; Mau, C.J.D., and West, C.A. (1994) *Proc. Natl. Acad. Sci. USA* **91**:8497-8501).

Each of the sense primers, A (SEQ ID NO:7), B (SEQ ID NO:8) and D (SEQ ID NO:10), was used for PCR in combination with antisense primer C (SEQ ID NO:9) by employing a broad range of amplification conditions. PCR was performed in a total volume of 50 µl containing 20 mM Tris/HCl (tris(hydroxymethyl) aminomethane/HCl, pH 8.4), 50 mM KCl, 5 mM $MgCl_2$, 200 µM of each dNTP, 1-5 µM of each primer, 2.5 units of *Taq* polymerase (BRL) and 5 µl of purified Grand fir stem cDNA library phage as template (1.5×10^9 pfu/ml). Analysis of the PCR reaction products by agarose gel electrophoresis (Sambrock, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) revealed that only the combination of primers C and D generated a specific PCR product of approximately 110 bps (base pairs). This PCR product was gel purified, ligated into pT7Blue (Novagen), and transformed into *E. coli* XL1-Blue cells. Plasmid DNA was prepared from 41 individual transformants and the inserts were sequenced (DyeDeoxy Terminator Cycle Sequencing, Applied Biosystems). Four different insert sequences were identified, and were designated as probes 1 (SEQ ID NO:11), 2 (SEQ ID NO:12), 4 (SEQ ID NO:13) and 5 (SEQ ID NO:14).

Subsequent isolation of four new cDNA species (*AG1.28* (SEQ ID NO:15); *AG2.2* (SEQ ID NO:1); *AG4.30* (SEQ ID NO:17) and *AG5.9* (SEQ ID NO:19)), encoding terpene synthases from Grand fir corresponding to probes 1 (SEQ ID NO:11), 2 (SEQ ID NO:12), 4 (SEQ ID NO:13) and 5 (SEQ ID NO:14), respectively, allowed the identification of three additional conserved sequence elements which were used to design a set of three new PCR primers: Primer E (5'-GGI GA(A/G) A(A/C)(A/G) (A/G)TI ATG GA(A/G) GA(A/G) GC-3')(SEQ ID NO:21); Primer F (5'-GA(A/G) (C/T)TI CA(G/A) (C/T)TI (A/C/T)(C/G/T)I (A/C)GI TGG TGG-3')(SEQ ID NO:22) and Primer G (5'-CCA (A/G)TT IA(A/G) ICC (C/T)TT IAC (A/G)TC-3')(SEQ ID NO:23).

Degenerate primer E (SEQ ID NO:21) was designed to conserved element GE(K/T)(V/I)M(E/D)EA (SEQ ID NO:26) and degenerate primer F (SEQ ID NO:22) was designed to conserved element Q(F/Y/D)(I/L)(T/L/R)RWW (SEQ ID NO:27) by comparing the sequences of five cloned terpene synthases from Grand fir: a monoterpene synthase corresponding to probe 2 (SEQ ID NO: 12), two sesquiterpene synthases corresponding to probe 4 (SEQ ID NO:13) and probe 5 (SEQ ID NO:14),

respectively, a previously described diterpene synthase (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268), and a truncated terpene synthase corresponding to probe 1 (SEQ ID NO:11). Degenerate primer G (SEQ ID NO:23) was designed according to the amino acid sequence DVIKG(F/L)NW (SEQ ID NO:28) obtained from a peptide generated by trypsin digestion of purified (-)-pinene synthase from Grand fir. Primers E (SEQ ID NO:21) and F (SEQ ID NO:22) were independently used for PCR amplification in combination with primer G (SEQ ID NO:23), with Grand fir stem cDNA library as template. The combination of primers E (SEQ ID NO:21) and G (SEQ ID NO:23) yielded a specific PCR product of approximately 1020 bps. This PCR product was ligated into pT7Blue and transformed into *E. coli* XL1-Blue. Plasmid DNA was prepared from 20 individual transformants and inserts were sequenced from both ends. The sequence of this 1022 bp insert was identical for all 20 plasmids and was designated as probe 3 (SEQ ID NO:24).

EXAMPLE 2

Screening a Wounded Grand fir Stem cDNA Library

For library screening, 100 ng of each probe was amplified by PCR, gel purified, randomly labeled with [α-^{32}P]dATP (Feinberg, A.P., and Vogelstein, B. (1984) *Anal. Biochem.* **137**:266-267), and used individually to screen replica filters of 10^5 plaques of the wound-induced Grand fir stem cDNA library plated on *E. coli* LE392. Hybridization with probes 1 (SEQ ID NO:11), 2 (SEQ ID NO:12), 4 (SEQ ID NO:13) and 5 (SEQ ID NO:14) was performed for 14 h at 65 °C in 3 x SSPE and 0.1% SDS. Filters were washed three times for 10 min at 55 °C in 3 x SSPE with 0.1% SDS and exposed for 12 h to Kodak XAR film at -70 °C. All of the λZAPII clones yielding positive signals were purified through a second round of hybridization (probe 1 (SEQ ID NO:11) gave 25 positives, probe 2 (SEQ ID NO:12) gave 16 positives, probe 4 (SEQ ID NO:13) gave 49 positives and probe 5 (SEQ ID NO:14) gave 12 positives).

Hybridization with probe 3 (SEQ ID NO:24) was performed as before, but the filters were washed three times for 10 min at 65 °C in 3 x SSPE and 0.1% SDS before exposure. Approximately 400 λZAPII clones yielded strong positive signals, and 34 of these were purified through a second round of hybridization at 65°C. Approximately 400 additional clones yielded weak positive signals with probe 3 (SEQ ID NO:24), and 18 of these were purified through a second round of hybridization for 20 h at 45 °C. Purified λZAP II clones isolated using all five

probes were *in vivo* excised as Bluescript II SK(-) phagemids and transformed into *E. coli* XLOLR according to the manufacturer's instructions (Stratagene). The size of each cDNA insert was determined by PCR using T3 (SEQ ID NO:29) and T7 (SEQ ID NO:30) promoter primers and selected inserts (>1.5 kb) were partially sequenced from both ends.

EXAMPLE 3

Grand fir Monoterpene Synthase cDNA Expression in E. coli and Enzyme Assays

Except for cDNA clones AG3.18 (SEQ ID NO:3) and AG3.48 (SEQ ID NO:31), all of the partially sequenced inserts were either truncated at the 5'-end, or were out of frame, or bore premature stop codons upstream of the presumptive methionine start codon. For the purpose of functional expression, a 2023 bp insert fragment, extending from nucleotides 75 to 2097 of the sequence set forth in SEQ ID NO:1, and a 1911 bp insert fragment, extending from nucleotide 1 to nucleotide 1910 of the sequence set forth in SEQ ID NO:3, were subcloned in frame into pGEX vectors (Pharmacia). A 2016 bp fragment extending from nucleotide 73 to nucleotide 2088 of the sequence set forth in SEQ ID NO:5 was subcloned in frame into the pSBETa vector (Schenk, P.M., Baumann, S., Mattes, R., and Steinbiss, H.-H. (1995) *Biotechniques* **19**, 196-200). To introduce suitable restriction sites for subcloning, fragments were amplified by PCR using primer combinations 2.2-*Bam*HI (5'-CAA AGG GAT CCA GAA TGG CTC TGG-3')(SEQ ID NO:33) and 2.2-*Not*I (5'-AGT AAG CGG CCG CTT TTT AAT CAT ACC CAC-3')(SEQ ID NO:34) with pAG2.2 insert (SEQ ID NO:1) as template, 3.18-*Eco*RI (5'-CTG CAG GAA TTC GGC ACG AGC-3')(SEQ ID NO:35) and 3.18-*Sma*I (5'-CAT AGC CCC GGG CAT AGA TTT GAG CTG-3')(SEQ ID NO:36) with pAG3.18 insert (SEQ ID NO:3) as template, and 10-*Nde*I (5-GGC AGG AAC ATA TGG CTC TCC TTT CTA TCG-3')(SEQ ID NO:37) and 10-*Bam*HI (5'-TCT AGA ACT AGT GGATCC CCC GGG CTG CAG-3')(SEQ ID NO:38) with pAG10 insert (SEQ ID NO:5) as template.

PCR reactions were performed in volumes of 50 µl containing 20 mM Tris/HCl (pH 8.8), 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100, 5 µg bovine serum albumin (BSA), 200 µM of each dNTP, 0.1 µM of each primer, 2.5 units of recombinant *Pfu* polymerase (Stratagene) and 100 ng plasmid DNA with the following program: denaturation at 94 °C, 1 min; annealing at 60 °C, 1 min; extension at 72 °C, 3.5 min; 35 cycles with final extension at 72 °C, 5 min. The PCR products were purified by agarose gel electrophoresis and used as

template for a secondary PCR amplification with the identical conditions in total volumes of 250 µl each. Products from this secondary amplification were digested with the above indicated restriction enzymes, purified by ultrafiltration and then ligated, respectively, into *Bam*HI/*Not*I-digested pGEX-4T-2 to yield plasmid pGAG2.2, into *Eco*RI/*Sma*I-digested pGEX-4T-3 to yield plasmid pGAG3.18, and into *Nde*I/*Bam*HI-digested pSBETa to yield plasmid pSBAG10; these plasmids were then transformed into *E. coli* XL1-Blue or *E. coli* BL21(DE3).

For expression, bacterial strains *E. coli* XLOLR/pAG3.18, *E. coli* XLOLR/pAG3.48, *E. coli* XL1-Blue/pGAG2.2, *E. coli* XL1-Blue/pGAG3.18, and *E. coli* BL21(DE3)/pSBAG10 were grown to A_{600} = 0.5 at 37 °C in 5 ml of LB medium (Sambrock, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) supplemented with 100 µg ampicillin/ml or 30 µg kanamycin/ml as determined by the vector. Cultures were then induced by addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside and grown for another 12 h at 20 °C. Cells were harvested by centrifugation (2000 x *g*, 10 min) and resuspended in either 1 ml monoterpene synthase assay buffer [50 mM Tris/HCl (pH 7.5), 500 mM KCl, 1 mM $MnCl_2$, 5 mM dithiothreitol, 0.05% (w/v) $NaHSO_3$ and 10% (v/v) glycerol], 1 ml sesquiterpene synthase assay buffer [10 mM dibasic potassium phosphate, 1.8 mM monobasic potassium phosphate (pH 7.3), 140 mM NaCl, 10 mM $MgCl_2$, 5 mM dithiothreitol, 0.05% (w/v) $NaHSO_3$ and 10% (v/v) glycerol], or 1 ml diterpene synthase assay buffer [30 mM Hepes (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid, pH 7.2), 7.5 mM $MgCl_2$, 5 mM dithiothreitol, 10 µM $MnCl_2$, 0.05% (w/v) $NaHSO_3$ and 10% (v/v) glycerol].

Cells were disrupted by sonication (Braun-Sonic 2000 with microprobe at maximum power for 15 seconds at 0-4°C), the homogenates were cleared by centrifugation (18,000 x *g*, 10 min), and 1 ml of the resulting supernatant was assayed for monoterpene synthase activity with 2.5 µM of [1-^{3}H]geranyl diphosphate, for sesquiterpene synthase activity with 3.5 µM [1-^{3}H]farnesyl diphosphate, or for diterpene synthase activity with 5 µM [1-^{3}H]geranylgeranyl diphosphate following standard protocols (Croteau, R., and Cane, D.E. (1985) *Methods Enzymol.* **110**:383-405; LaFever, R.E., Stofer Vogel, B., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **313**:139-149; Dehal, S.S., and Croteau, R. (1988) *Arch. Biochem. Biophys.* **261**:346-356). In the case of the monoterpene synthase and sesquiterpene synthase assays, the incubation mixture was overlaid with 1 ml pentane

to trap volatile products. In all cases, after incubation at 31°C for 2 h, the reaction mixture was extracted with pentane (3 x 1 ml) and the combined extract was passed through a 1.5 ml column of anhydrous $MgSO_4$ and silica gel (Mallinckrodt 60 Å) to provide the terpene hydrocarbon fraction free of oxygenated metabolites. The columns were subsequently eluted with 3 x 1 ml of ether to collect any oxygenated products, and an aliquot of each fraction was taken for liquid scintillation counting to determine conversion rate.

Product Identification – To obtain sufficient product for analysis by radio-GLC (gas liquid chromatography), chiral capillary GLC and GLC-MS (mass spectrum/spectrometry), preparative-scale enzyme incubations were carried out. Thus, the enzyme was prepared from 50 ml of cultured bacterial cells by extraction with 3 ml of assay buffer as above, and the extracts were incubated with excess substrate overnight at 31°C. The hydrocarbon fraction was isolated by elution through $MgSO_4$-silica gel as before, and the pentane eluate was concentrated for evaluation by capillary radio-GLC as described (Croteau, R., and Satterwhite, D.M. (1990) *J. Chromatogr.* **500**:349-354), by chiral column capillary GLC (Lewinsohn, E., Savage, T.J., Gijzen, M., and Croteau, R. (1993) *Phytochem. Anal.* **4**:220-225), and by combined GLC-MS [Hewlett-Packard 6890 GC-MSD with cool (40°C) on-column injection, detection via electron impact ionization (70 eV), He carrier at 0.7 psi., column: 0.25 mm i.d. x 30 m fused silica with 0.25 µm film of 5MS (Hewlett-Packard) programmed from 35°C (5 min hold) to 230°C at 5°C/min].

EXAMPLE 4

Sequence Analysis

Inserts of all recombinant bluescript plasmids and pGEX plasmids were completely sequenced on both strands via primer walking and nested deletions (Sambrock, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) using the DyeDeoxy Terminator Cycle Sequencing method (Applied Biosystems). Sequence analysis was done using the Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI.

EXAMPLE 5

RNA Extraction and Northern Blotting

Grand fir sapling stem tissue was harvested prior to wounding or two days after wounding by a standard procedure (Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**:267-273). Total RNA was isolated (Lewinsohn,

E., Steele, C.L., and Croteau, R. (1994) *Plant Mol. Biol. Rep.* **12**:20-25) and 20 μg of RNA per gel lane was separated under denaturing conditions (Sambrock, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) and transferred to nitrocellulose membranes (Schleicher and Schuell) according to the manufacturer's protocol. To prepare hybridization probes, cDNA fragments of 1.4-1.5 kb were amplified by PCR from *AG2.2* (SEQ ID NO:1) with primer JB29 (5'-CTA CCA TTC CAA TAT CTG-3')(SEQ ID NO:39) and primer 2-8 (5'-GTT GGA TCT TAG AAG TTC CC-3')(SEQ ID NO:40), from *AG3.18* (SEQ ID NO:3) with primer 3-9 (5'-TTT CCA TTC CAA CCT CTG GG-3')(SEQ ID NO:41) and primer 3-11 (5'-CGT AAT GGA AAG CTC TGG CG-3')(SEQ ID NO:42), and from *AG10* (SEQ ID NO:5) with primer 7-1 (5'-CCT TAC ACG CCT TTG GAT GG-3')(SEQ ID NO:43) and primer 7-3 (5'-TCT GTT GAT CCA GGA TGG TC-3')(SEQ ID NO:44). The probes were randomly labeled with [α-^{32}P]dATP (Feinberg, A.P., and Vogelstein, B. (1984) *Anal Biochem.* **137**:266-267). Blots were hybridized for 24 h at 55°C in 3 x SSPE and 0.1% SDS, washed at 55°C in 1 x SSPE and 0.1%, SDS and subjected to autoradiography as described above at -80°C for 24 h.

EXAMPLE 6

Cloning and Characterization of Clones AG1.28 (SEQ ID NO:15), AG2.2 (SEQ ID NO:1), AG4.30 (SEQ ID NO:17) and AG 5.9 (SEQ ID NO:19)

Similarity-Based Cloning of Grand fir Terpene Synthases – Grand fir has been developed as a model system for the study of induced oleoresin production in conifers in response to wounding and insect attack (Johnson, M.A., and Croteau, R. (1987) in *Ecology and Metabolism of Plant Lipids* (Fuller, G., and Nes, W.D., eds) pp. 76-91, American Chemical Society Symposium Series 325, Washington, DC; Gijzen, M., Lewinsohn, E., Savage, T.J., and Croteau, R.B. (1993) in *Bioactive Volatile Compounds from Plants* (Teranishi, R., Buttery, R.G., and Sugisawa, H., eds) pp. 8-22, American Chemical Society Symposium Series 525, Washington, DC; Raffa, K. F., and Berryman, A.A. (1982) *Can. Entomol.* **114**:797-810; Steele, C., Lewinsohn, E., and Croteau, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**:4164-4168; Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) in *Regulation of Isopentenoid Metabolism* (Nes, W.D., Parish, E.J., and Trzaskos, J.M., eds) pp. 8-17, American Chemical Society Symposium Series 497, Washington, DC). The chemistry and biosynthesis of the oleoresin monoterpenes, sesquiterpenes and diterpenes have been well defined (Lewinsohn, E., Savage, T.J., Gijzen, M., and

Croteau, R. (1993) *Phytochem. Anal.* **4**:220-225; Lewinsohn, E., Gijzen, M., and Croteau, R. (1991) *Plant Physiol.* **96**:44-49; Funk, C., Lewinsohn, E., Stofer Vogel, B., Steele C., and Croteau, R. (1994) *Plant Physiol.* **106**:999-1005; Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**:267-273; Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173; LaFever, R.E., Stofer Vogel, B., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **313**:139-149; Funk, C., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **308**:258-266); however, structural analysis of the responsible terpene synthases as well as studies on the regulation of oleoresinosis require the isolation of cDNA species encoding the terpene synthases. Protein purification from conifers, as the basis for cDNA isolation, has been of limited success (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268) and thus far has not permitted cloning of any of the monoterpene synthases from these species (Steele, C., Lewinsohn, E., and Croteau, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**:4164-4168).

As a possible alternative to protein-based cloning of terpene synthases, a homology-based PCR strategy was proposed (Steele, C., Lewinsohn, E., and Croteau, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**:4164-4168) that was founded upon the three terpene synthases of plant origin then available, a monoterpene synthase, (-)-4*S*-limonene synthase, from spearmint (*Mentha spicata*, Lamiaceae) (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024), a sesquiterpene synthase, 5-*epi*-aristolochene synthase, from tobacco (*Nicotiana tabacum*, Solanaceae) (Facchini, P.J., and Chappell, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**:11088-11092), and a diterpene synthase, casbene synthase, from castor bean (*Ricinus communis*, Euphorbiaceae) (Mau, C.J.D., and West, C.A. (1994) *Proc. Natl. Acad. Sci. USA* **91**:8497-8501). Despite the taxonomic distances between these three angiosperm species and the differences in substrate utilized, reaction mechanism and product type of the three enzymes, a comparison of the deduced amino acid sequences identified several conserved regions that appeared to be useful for the design of degenerate PCR primers (see Example 1). Using cDNA from a wound-induced Grand fir stem library as template, PCR primers C (SEQ ID NO:9) and D (SEQ ID NO:10) amplified products corresponding to four distinct sequence groups, all of which showed significant similarity to sequences of cloned terpene synthases of plant origin. The four different inserts were designated as probes 1 (SEQ ID NO:11), 2 (SEQ ID NO:12), 4

(SEQ ID NO:13) and 5 (SEQ ID NO:14), and were employed for isolation of the corresponding cDNA clones by plaque hybridization.

Screening of 10^5 cDNA phage plaques from the wounded Grand fir stem library, with each of the four probes, yielded a four-fold difference in the number of positives, most likely reflecting different levels of expression of the corresponding genes. Size selected inserts (>1.5 kb) of purified and *in vivo* excised clones were partially sequenced from both ends, and were shown to segregate into four distinct groups corresponding to the four hybridization probes. Since all cDNAs corresponding to probes 1 (SEQ ID NO:11), 4 (SEQ ID NO:13) and 5 (SEQ ID NO:14) were truncated at their 5'-ends, only inserts of the largest representatives of each group, clone *AG1.28* (SEQ ID NO:15), clone *AG2.2* (SEQ ID NO:1) (apparently full length), clone *AG4.30* (SEQ ID NO:17) and clone *AG5.9* (SEQ ID NO:19), were completely sequenced. Clone *AG1.28* (SEQ ID NO:15)(2424 bps) includes an open reading frame (ORF) of 2350 nucleotides (nts) encoding 782 amino acids (SEQ ID NO:16); clone *AG2.2* (SEQ ID NO:1)(2196 bps), includes an ORF of 1881 nts encoding 627 amino acids (SEQ ID NO:2); clone *AG4.30* (SEQ ID NO:17)(1967 bps) includes an ORF of 1731 nts encoding 577 amino acids (SEQ ID NO:18) and clone *AG5.9* (SEQ ID NO:19)(1416 bps) includes an ORF of 1194 nucleotides encoding 398 amino acids (SEQ ID NO:20).

cDNA clones *AG1.28* (SEQ ID NO:15), *AG2.2* (SEQ ID NO:1), *AG4.30* (SEQ ID NO:17) and *AG5.9* (SEQ ID NO:19) were compared pairwise with each other and with other cloned plant terpene synthases. Truncated clone *AG1.28* (SEQ ID NO:15) resembled most closely in size and sequence (72% similarity, 49% identity) a diterpene cyclase, abietadiene synthase, from Grand fir (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268). Clones *AG4.30* (SEQ ID NO:17) and *AG5.9* (SEQ ID NO:19) share approximately 80% similarity (60% identity) at the amino acid level, and are almost equally distant from both clone *AG1.28* (SEQ ID NO:15) and full-length clone *AG2.2* (SEQ ID NO:1)(range of 65-70% similarity and 45-47% identity); the amino acid sequence similarity between *AG1.28* (SEQ ID NO:15) and *AG2.2* (SEQ ID NO:1) is 65% (41% identity). Considering the high level of homology between *AG4.30* (SEQ ID NO:17) and *AG5.9* (SEQ ID NO:19), these comparisons suggest that the four new cDNAs, *AG1.28* (SEQ ID NO:15), *AG2.2* (SEQ ID NO:1), *AG4.30* (SEQ ID NO:17) and *AG5.9* (SEQ ID NO:19), represent the three major subfamilies

of Grand fir terpene synthase genes encoding monoterpene synthases, sesquiterpene synthases and diterpene synthases.

Identification of cDNA Clone AG2.2 (SEQ ID NO:1) ***as Myrcene Synthase*** – The pAG2.2 insert (SEQ ID NO:1) appeared to be a full-length clone encoding a protein of molecular weight 72,478 with a calculated pI at 6.5. The size of the translated protein encoded by *AG2.2* (SEQ ID NO:1) (627 residues)(SEQ ID NO:2) is in the range of the monoterpene synthase preproteins for limonene synthase from spearmint (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024) and *Perilla frutescens* (Yuba, A., Yazaki, K., Tabata, M., Honda, G., and Croteau, R. (1996) *Arch. Biochem. Biophys.* **332**:280-287), but is about 240 amino acids shorter than the two gymnosperm diterpene synthase preproteins for abietadiene synthase (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268) and taxadiene synthase (Wildung, M.R., and Croteau, R. (1996) *J. Biol. Chem.* **271**:9201-9204). Monoterpene and diterpene biosynthesis are compartmentalized in plastids whereas sesquiterpene biosynthesis is cytosolic (reviewed in Kleinig, H. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**:39-53; Gershenzon, J., and Croteau, R. (1993) in *Lipid Metabolism in Plants* (Moore, T.S. Jr., ed) pp. 339-388, CRC Press, Boca Raton, FL; McGarvey, D.J., and Croteau, R. (1995) *Plant Cell* 7, 1015-1026); thus, monoterpene and diterpene synthases are encoded as preproteins bearing an amino-terminal transit peptide for import of these nuclear gene products into plastids where they are proteolytically processed to the mature forms (Keegstra, K., Olsen, J.J., and Theg, S.M. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**:471-501; von Heijne, G., Stepphuhn, J., and Herrmann (1989) *Eur. J. Biochem.* **180**:535-545). Both the size of the deduced protein and the presence of an *N*-terminal domain (of 60 to 70 amino acids) with features characteristic of a targeting sequence [rich in serine residues (16-18%) and low in acidic residues (four Asp or Glu) (Keegstra, K., Olsen, J.J., and Theg, S.M. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**:471-501; von Heijne, G., Stepphuhn, J., and Herrmann (1989) *Eur. J. Biochem.* **180**:535-545)] suggest that *AG2.2* (SEQ ID NO:1) encodes a monoterpene synthase rather than a sesquiterpene synthase or a diterpene synthase.

Since pAG2.2 contained the terpene synthase insert in reversed orientation, the ORF was subcloned in frame with glutathione *S*-transferase, for ultimate ease of purification (Bohlmann, J., DeLuca, V., Eilert, U., and Martin, W. (1995) *Plant J.* 7:491-501; Bohlmann, J., Lins, T., Martin, W., and Eilert, U. (1996) *Plant Physiol.*

111:507-514), into pGEX-4T-2, yielding plasmid pGAG2.2. The recombinant fusion protein was expressed in *E. coli* strain XL1-Blue/pGAG2.2, then extracted and assayed for monoterpene synthase, sesquiterpene synthase and diterpene synthase activity using tritium labeled geranyl diphosphate, farnesyl diphosphate and geranylgeranyl diphosphate as the respective substrate. Enzymatic production of a terpene olefin was observed only with geranyl diphosphate as substrate, and the only product was shown to be myrcene by radio-GLC and GLC-MS comparison to an authentic standard (FIGURE 3). Bacteria transformed with pGEX vector containing the *AG2.2* insert (SEQ ID NO:1) in antisense orientation did not afford detectable myrcene synthase activity when induced, and the protein isolated and assayed as above. A myrcene synthase cDNA has not been obtained previously from any source, although myrcene is a minor co-product (2%) of the native and recombinant limonene synthase from spearmint (Rajaonarivony, J.I.M., Gershenzon, J., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **296**:49-57; Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024) and of several enzymes from sage (Croteau, R., and Satterwhite, D.M. (1989) *J. Biol. Chem.* **264**:15309-15315). cDNA cloning and functional expression of myrcene synthase, which is one of several wound-inducible monoterpene synthase activities of Grand fir (Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**:267-273), demonstrates that this acyclic monoterpene is formed by a distinct enzyme and is not a co-product of another synthase.

EXAMPLE 7

Cloning and Characterization of Clones AG3.18 (SEQ ID NO:3) Encoding (-)-Pinene Synthase and cDNA Clone AG10 (SEQ ID NO:5) Encoding (-)-Limonene Synthase

Identification of cDNA Clone AG3.18* (SEQ ID NO:3) *as (-)-Pinene Synthase and cDNA Clone AG10* (SEQ ID NO:5) *as (-)-Limonene Synthase – Alignment of the four new terpene synthase cDNA sequences (*AG1.28* (SEQ ID NO:15), *AG2.2* (SEQ ID NO:1), *AG4.30* (SEQ ID NO:17) and *AG.5.9* (SEQ ID NO:19)), and that for abietadiene synthase (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268), allowed the identification of several conserved sequence motifs among this enzyme family from Grand fir, which provided the foundation for an extended similarity-based cloning approach. Two new sense primers E (SEQ ID NO:21) and F (SEQ ID NO:22) were designed according to conserved sequence elements, whereas a degenerate antisense

primer G (SEQ ID NO:23) was designed based upon very limited amino acid sequence information from pinene synthase (see Example 1). Only the combination of primers E (SEQ ID NO:21) and G (SEQ ID NO:23) amplified a specific product of 533 bps, which was designated as probe 3 (SEQ ID NO:24).

Hybridization of 10^5 Grand fir λZAP II cDNA clones with probe 3 (SEQ ID NO:24) yielded two types of signals comprised of about 400 strongly positive clones and an equal number of weak positives, indicating that the probe recognized more than one type of cDNA. Thirty-four of the former clones and 18 of the latter were purified, the inserts were selected by size (2.0-2.5 kb), and the *in vivo* excised clones were partially sequenced from both ends. Those clones which afforded weak hybridization signals were shown to contain inserts that were either identical to myrcene synthase clone *AG2.2* (SEQ ID NO:1) or exhibited no significant sequence similarity to terpene synthases. Clone AG3.48 (SEQ ID NO:31) contained the myrcene synthase ORF in the correct orientation and in frame for expression from the Bluescript plasmid vector. This cDNA was functionally expressed in *E. coli* and the resulting enzyme was shown to accept only geranyl diphosphate as the prenyl diphosphate substrate and to produce myrcene as the exclusive reaction product. This finding with AG3.48 (SEQ ID NO:31) confirms that expression of AG2.2 (SEQ ID NO:1) as the glutathione *S*-transferase fusion protein from pGAG2.2 does not influence substrate utilization or product outcome of the myrcene synthase.

Clones which gave strong hybridization signals segregated into distinct sequence groups represented by clone *AG3.18* (SEQ ID NO:3)(2018 bp insert with ORF of 1884 nt; encoded protein of 628 residues at 71,505 Da and pI of 5.5) and *AG10* (SEQ ID NO:5)(2089 bp insert with ORF of 1911 nt; encoded protein of 637 residues at 73,477 Da and pI of 6.4). *AG3.18* (SEQ ID NO:3) and *AG10* (SEQ ID NO:5) form a subfamily together with the myrcene synthase clone *AG2.2* (SEQ ID NO:1) that is characterized by a minimum of 79% pairwise similarity (64% identity) at the amino acid level. Like myrcene synthase, both *AG3.18* (SEQ ID NO:3) and *AG10* (SEQ ID NO:5) encode *N*-terminal sequences of 60 to 70 amino acids which are rich in serine (19-22% and 11-15%, respectively) and low in acidic residues (4 and 2, respectively) characteristic of plastid transit peptides (Keegstra, K., Olsen, J.J., and Theg, S.M. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**:471-501; von Heijne, G., Stepphuhn, J., and Herrmann (1989) *Eur. J. Biochem.* **180**:535-545).

Plasmid pAG3.18 (SEQ ID NO:3) contained the presumptive terpene synthase ORF in frame for direct expression from the bluescript plasmid, whereas the

AG10 (SEQ ID NO:5) ORF was in reversed orientation. Both *AG3.18* (SEQ ID NO:3) and *AG10* (SEQ ID NO:5) were subcloned into expression vectors yielding plasmids pGAG3.18 and pSBAG10. Recombinant proteins were expressed in bacterial strain *E. coli* XLOLR/pAG3.18, *E. coli* XL1-Blue/pGAG3.18 and *E. coli* BL21(DE3)/pSBAG10. When extracts of the induced cells were tested for terpene synthase activity with all of the potential prenyl diphosphate substrates, only geranyl diphosphate was utilized. Extracts from *E. coli* BL21(DE3)/pSBAG10 converted geranyl diphosphate to limonene as the major product with lesser amounts of α-pinene, β-pinene and β-phellandrene, as determined by radio-GLC and combined GLC-MS (FIGURE 5). Chiral phase capillary GLC on β-cyclodextrin revealed the limonene product to be the (-)-4*S*-enantiomer and the pinene products to be the related (-)-(1*S*:5*S*)-enantiomers. Although optically pure standards were not available for the analysis, stereochemical considerations suggest that the minor product β-phellandrene is also the mechanistically related (-)-(4*S*)-antipode (Gambliel, H., and Croteau, R. (1984) *J. Biol. Chem.* **259**:740-748; Croteau, R., Satterwhite, D.M., Cane, D.E., and Chang, C.C. (1988) *J. Biol. Chem.* **263**:10063-10071; Wagschal, K., Savage, T.J., and Croteau, R. (1991) *Tetrahedron* **47**:5933-5944; Croteau, R., Satterwhite, D.M., Wheeler, C.J., and Felton, N.M. (1989) *J. Biol. Chem.* **264**:2075-2080; LaFever, R.E., and Croteau, R. (1993) *Arch. Biochem. Biophys.* **301**:361-366). Similar analysis of the monoterpene products generated from geranyl diphosphate by cell-free extracts of *E. coli* XLOLR/pAG3.18 and *E. coli* XL1-Blue/pGAG3.18 demonstrated the presence of a 42:58% mixture of α-pinene and β-pinene (FIGURE 4), the same product ratio previously described for the purified, native (-)-pinene synthase from Grand fir (Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173). Chiral phase capillary GLC confirmed the products of the recombinant pinene synthase to be the (-)-(1*S*:5*S*)-enantiomers, as expected. No other monoterpene co-products were detected with the recombinant (-)-pinene synthase, as observed previously for the native enzyme (Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173).

Evidence for the formation of both α- and β-pinene by a single enzyme has been previously provided through co-purification studies, and differential inhibition and inactivation studies, as well as by isotopically sensitive branching experiments (Gambliel, H., and Croteau, R. (1984) *J. Biol. Chem.* **259**:740-748; Wagschal, K.C., Pyun, H.-J., Coates, R.M., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **308**:477-

487; Wagschal, K., Savage, T.J., and Croteau, R. (1991) *Tetrahedron* **47**:5933-5944; Croteau, R., Wheeler, C.J., Cane, D.E., Ebert, R., and Ha, H.-J. (1987) *Biochemistry* **26**:5383-5389). The cDNA cloning of pinene synthase provides the ultimate proof that a single enzyme forms both products. The calculated molecular weight of the (-)-pinene synthase deduced from *AG3.18* (SEQ ID NO:3) is approximately 64,000 (excluding the putative transit peptide), which agrees well with the molecular weight of 63,000 established for the native enzyme from Grand fir by gel permeation chromatography and SDS-PAGE (Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173).

A limonene synthase cDNA has thus far been cloned only from two very closely related angiosperm species (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024; Yuba, A., Yazaki, K., Tabata, M., Honda, G., and Croteau, R. (1996) *Arch. Biochem. Biophys.* **332**:280-287), and the isolation of a pinene synthase cDNA has not been reported before. Pinene synthase has previously received considerable attention as a major defense-related monoterpene synthase in conifers (Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**:267-273; Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173). In the Grand fir cDNA library, which was synthesized from mRNA obtained from wound-induced sapling stems, clones corresponding to pinene synthase are at least ten times more abundant than clones for myrcene synthase. This finding reflects the relative proportions of the induced levels of activities of these enzymes in Grand fir saplings; pinene synthase and limonene synthase are the major monoterpene synthase activities whereas the induced level of myrcene synthase activity is relatively low (Gijzen, M., Lewinsohn, E., and Croteau, R. (1991) *Arch. Biochem. Biophys.* **289**:267-273). The cDNAs for inducible monoterpene synthases provide probes for genetic and molecular analysis of oleoresin-based defense in conifers. Northern blots (FIGURE 6) of total RNA extracted from non-wounded sapling stems and from stems two days after wounding (when enzyme activity first appears) were probed with cDNA fragments for *AG2.2* (SEQ ID NO:1), *AG3.18* (SEQ ID NO:3) and *AG10* (SEQ ID NO:5), and thus demonstrated that increased mRNA accumulation for monoterpene synthases is responsible for this induced, defensive response in Grand fir. The availability of cloned, defense-related monoterpene synthases presents several possible avenues for transgenic manipulation of oleoresin composition to improve tree resistance to bark beetles and other pests. For example, altering the

monoterpene content of oleoresin may chemically disguise the host and decrease insect aggregation by changing the levels of pheromone precursors or predator attractants, or lower infestation by increasing toxicity toward beetles and their pathogenic fungal associates (Johnson, M.A., and Croteau, R. (1987) in *Ecology and Metabolism of Plant Lipids* (Fuller, G., and Nes, W.D., eds) pp. 76-91, American Chemical Society Symposium Series 325, Washington, DC; Gijzen, M., Lewinsohn, E., Savage, T.J., and Croteau, R.B. (1993) in *Bioactive Volatile Compounds from Plants* (Teranishi, R., Buttery, R.G., and Sugisawa, H., eds) pp. 8-22, American Chemical Society Symposium Series 525, Washington, DC; Byers, J.A. (1995) in *Chemical Ecology of Insects 2* (Cardé, R.T., and Bell, W.J., eds) pp. 154-213, Chapman and Hall, New York).

EXAMPLE 8

Properties of the Recombinant Monoterpene Synthases Encoded by cDNA Clones *AG2.2 (SEQ ID NO:1), AG3.18 (SEQ ID NO:3)* and *AG10 (SEQ ID NO:5)*

All three recombinant enzymes require Mn^{2+} for activity, and Mg^{2+} is essentially ineffective as the divalent metal ion cofactor. This finding confirms earlier results obtained with the native monoterpene synthases of Grand fir and lodgepole pine (*Pinus contorta*) (Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173; Savage, T.J., Hatch, M.W., and Croteau, R. (1994) *J. Biol. Chem.* **269**:4012-4020). All terpene synthases and prenyltransferases are thought to employ a divalent metal ion, usually Mg^{2+} or Mn^{2+}, in the ionization steps of the reaction sequence to neutralize the negative charge of the diphosphate leaving group (Croteau, R. (1987) *Chem. Rev.* **87**:929-954; Cane, D.E. (1992) *Ciba Found. Symp. Ser.* **171**:163-167; Poulter, C.D., and Rilling, H.C. (1981) in *Biosynthesis of Isoprenoid Compounds* (Porter, J.W., and Spurgeon, S.L., eds) Vol. 1, pp. 161-224, Wiley & Sons, New York), and all relevant sequences thus far obtained bear a conserved aspartate rich element (DDXXD)(SEQ ID NO:45) considered to be involved in divalent metal ion binding (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268; Ashby, M.N., and Edwards, P.A. (1990) *J. Biol. Chem.* **265**:13157-13164; Chen, A., Kroon, P.A., and Poulter, D.C. (1994) *Protein Sci.* **3**:600-607; Tarshis, L.C., Yan, M., Poulter, C.D., and Sacchettini, J.C. (1994) *Biochemistry* **33**:10871-10877; Cane, D.E., Sohng, J.K., Lamberson, C.R., Rudnicki, S.M., Wu, Z., Lloyd, M.D., Oliver, J.S., and Hubbard, B.R. (1994) *Biochemistry* **33**:5846-5857; Proctor, R.H., and

Hohn, T.M. (1993) *J. Biol. Chem.* **268**:4543-4548). In addition to this strict, general dependence on a divalent metal ion, the monoterpene synthases of conifers are unique in their further requirement for a monovalent cation (K^+), a feature that distinguishes the gymnosperm monoterpene synthases from their counterparts from angiosperm species and implies a fundamental structural and/or mechanistic difference between these two families of catalysts (Savage, T.J., Hatch, M.W., and Croteau, R. (1994) *J. Biol. Chem.* **269**:4012-4020). All three recombinant monoterpene synthases depend upon K^+, with maximum activity achieved at approximately 500 mM KCl. A requirement for K^+ has been reported for a number of different types of enzymes, including those that catalyze phosphoryl cleavage or transfer reactions (Suelter, C.H. (1970) *Science* **168**:789-794) such as Hsc70 ATPase (Wilbanks, S.M., and McKay, D.B. (1995) *J. Biol. Chem.* **270**:2251-2257). The crystal structure of bovine Hsc70 ATPase indicates that both Mg^{2+} and K^+ interact directly with phosphate groups of the substrate and implicates three active site aspartate residues in Mg^{2+} and K^+ binding (Wilbanks, S.M., and McKay, D.B. (1995) *J. Biol. Chem.* **270**:2251-2257), reminiscent of the proposed role of the conserved DDXXD (SEQ ID NO:45) motif of the terpene synthases and prenyltransferases in divalent cation binding, a function also supported by recent site directed mutagenesis (Marrero, P.F., Poulter, C.D., and Edwards, P.A. (1992) *J. Biol. Chem.* **267**:21873-21878; Joly, A., and Edwards, P.A. (1993) *J. Biol. Chem.* **268**: 26983-26989; Song, L., and Poulter, C.D. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**:3044-3048; Koyama, T., Tajima, M., Sano, H., Doi, T., Koike-Takeshita, A., Obata, S., Nishino, T., and Ogura, K. (1996) *Biochemistry* **35**:9533-9538) and by X-ray structural analysis (Tarshis, L.C., Yan, M., Poulter, C.D., and Sacchettini, J.C. (1994) *Biochemistry* **33**:10871-10877) of farnesyl diphosphate synthase.

cDNA cloning and functional expression of the myrcene, limonene and pinene synthases from Grand fir represent the first example of the isolation of multiple synthase genes from the same species, and provide tools for evaluation of structure-function relationships in the construction of acyclic, monocyclic and bicyclic monoterpene products and for detailed comparison to catalysts from phylogenetically distant plants that carry out ostensibly identical reactions (Gambliel, H., and Croteau, R. (1984) *J. Biol. Chem.* **259**:740-748; Rajaonarivony, J.I.M., Gershenzon, J., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **296**:49-57; Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024; Adam, K.-P., Crock, J., and Croteau, R. (1996) *Arch.*

Biochem. Biophys. **332**:352-356). The recent acquisition of cDNA isolates encoding sesquiterpene synthases and diterpene synthases (Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268) from Grand fir should, together with the monoterpene synthases, also permit addressing the structural basis of chain-length specificity for prenyl diphosphate substrates in this family of related enzymes.

EXAMPLE 9

Sequence Comparison of Cloned Monoterpene Synthases

Previous studies based on substrate protection from inactivation with selective amino acid modifying reagents have implicated functionally important cysteine, histidine and arginine residues in a range of different monoterpene synthases (Rajaonarivony, J.I.M., Gershenzon, J., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **296**:49-57; Lewinsohn, E., Gijzen, M., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **293**:167-173; Savage, T.J., Hatch, M.W., and Croteau, R. (1994) *J. Biol. Chem.* **269**:4012-4020; Rajaonarivony, J.I.M., Gershenzon, J., Miyazaki, J., and Croteau, R. (1992) *Arch. Biochem. Biophys.* **299**:77-82; Savage, T.J., Ichii, H., Hume, S.D., Little, D.B., and Croteau, R. (1994) *Arch. Biochem. Biophys.* **320**:257-265). Sequence alignment of 21 terpene synthases of plant origin (Colby, S.M., Alonso, W.R., Katahira, E.J., McGarvey, D.J., and Croteau, R. (1993) *J. Biol. Chem.* **268**:23016-23024; Stofer Vogel, B., Wildung, M.R., Vogel, G., and Croteau, R. (1996) *J. Biol. Chem.* **271**:23262-23268; Facchini, P.J., and Chappell, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**:11088-11092; Mau, C.J.D., and West, C.A. (1994) *Proc. Natl. Acad. Sci. USA* **91**:8497-8501; Yuba, A., Yazaki, K., Tabata, M., Honda, G., and Croteau, R. (1996) *Arch. Biochem. Biophys.* **332**:280-287; Wildung, M.R., and Croteau, R. (1996) *J. Biol. Chem.* **271**:9201-9204; Yamaguchi, S., Saito, T., Abe, H., Yamane, H., Murofushi, N., and Kamiya, Y. (1996) *Plant J.* **10**:203-213; Dudareva, N., Cseke, L., Blanc, V.M., and Pichersky, E. (1996) *Plant Cell* **8**:1137-1148; Chen, X.-Y., Chen, Y., Heinstein, P., and Davisson, V.J. (1995) *Arch. Biochem. Biophys.* **324**:255-266; Chen, X.-Y., Wang, M., Chen, Y., Davisson, J., and Heinstein, P. (1996) *J. Nat. Prod.* **59**:944-951; Back, K., and Chappell, J. (1995) *J. Biol. Chem.* **270**:7375-7381) reveals two absolutely conserved arginine residues, corresponding to Arg^{184} and Arg^{365} of pinene synthase (SEQ ID NO:4), one highly conserved cysteine residue (pinene synthase Cys^{543})(SEQ ID NO:4), and one highly conserved histidine residue (pinene synthase His^{186})(SEQ ID NO:4). The DDXXD (SEQ ID NO:45) sequence motif (pinene synthase Asp^{379}, Asp^{380} and Asp^{383}) (SEQ

ID NO:4) is absolutely conserved in all relevant plant terpene synthases, as are several other amino acid residues corresponding to Phe^{198}, Leu^{248}, Glu^{322}, Trp^{329}, Trp^{460} and Pro^{467} of pinene synthase (SEQ ID NO:4).

Amino acid sequences of the plant terpene synthases were compared with each other and with the deduced sequences of several sesquiterpene synthases cloned from microorganisms (Proctor, R.H., and Hohn, T.M. (1993) *J. Biol. Chem.* **268**:4543-4548; Back, K., and Chappell, J. (1995) *J. Biol. Chem.* **270**:7375-7381; Hohn, T.M., and Desjardins, A.E. (1992) *Mol. Plant-Microbe Interactions* **5**:249-256). As with all other plant terpene synthases, no significant conservation in primary sequence exists between the monoterpene synthases from Grand fir and the terpene synthases of microbial origin, except for the DDXXD (SEQ ID NO:45) sequence motif previously identified as a common element of all terpene synthases, and prenyltransferases which employ a related electrophilic reaction mechanism (Croteau, R., Wheeler, C.J., Cane, D.E., Ebert, R., and Ha, H.-J. (1987) *Biochemistry* **26**:5383-5389; Chen, A., Kroon, P.A., and Poulter, D.C. (1994) *Protein Sci.* **3**:600-607; McCaskill, D., and Croteau, R. (1997) *Adv. Biochem. Engineering Biotech.* **55**:108-146). The evidence is presently insufficient to determine whether extant plant and microbial terpene synthases represent divergent evolution from a common ancestor, which may also have given rise to the prenyltransferases, or whether these similar catalysts evolved convergently.

EXAMPLE 10

Alteration of Monoterpene Levels and Composition in Plant Seeds

In accordance with the present invention, methods for increasing production of monoterpene compounds in a plant, particularly in plant seeds, are provided. The methods involve transforming a plant cell with a nucleic acid sequence encoding at least one gymnosperm monoterpene synthase, such as those encoded by the nucleic acid sequences set forth in SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:5. This has the effect of altering monoterpene biosynthesis, thereby increasing the production of monoterpenes, as well as providing novel seed oils having desirable monoterpene compositions. In this manner, the transformed seed provides a factory for the production of modified oils. The modified oil itself may be used and/or the compounds in the oils can be isolated. Thus, the present invention allows for the production of particular monoterpenes of interest as well as speciality oils.

The nucleic acid encoding the monoterpene synthases of the present invention can be used in expression cassettes for expression in the transformed plant tissues.

To alter the monoterpene levels in a plant of interest, the plant is transformed with at least one expression cassette comprising a transcriptional initiation region linked to a nucleic acid sequence encoding a monoterpene synthase. Such an expression cassette is provided with a plurality of restriction sites for insertion of the nucleic acid sequence encoding a monoterpene synthase so that it is under the transcriptional regulation of the regulatory regions.

The transcriptional initiation sequence may be native or analogous to the host or foreign or heterologous to the host. In this regard, the term "foreign" means that the transcriptional initiation sequence is not found in the wild-type host into which the transcriptional initiation region is introduced.

Of particular interest are those transcriptional initiation regions associated with storage proteins, such as napin, cruciferin, β-conglycinin, phaseolin, globulin or the like, and proteins involved in fatty acid biosynthesis, such as acyl carrier protein (ACP). See, U.S. Patent No. 5,420,034, herein incorporated by reference.

The transcriptional cassette will preferably include, in the 5' to 3' direction of transcription, a transcriptional and translational initiation region, a gymnosperm monoterpene synthase DNA sequence of interest, and a transcriptional and translational termination region functional in plants. The termination region may be from the same organism as the transcriptional initiation region, may be from the same organism as the monoterpene synthase DNA, or may be derived from another source. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. Other termination sequences are set forth in Guerineau et al., (1991), *Mol. Gen. Genet.*, **262**:141-144; Proudfoot, (1991), *Cell*, **64**:671-674; Sanfacon et al., (1991), *Genes Dev.*, **5**:141-149; Mogen et al., (1990), *Plant Cell*, **2**:1261-1272; Munroe et al., (1990), *Gene*, **91**:151-15 8; Ballas et al., (1989), *Nucleic Acids Res.*, **17**:7891-7903; Joshi et al., (1987), *Nucleic Acid Res.*, **15**:9627-9639).

In the presently preferred form of the invention, a nucleic acid sequence encoding a gymnosperm monoterpene synthase protein will be targeted to plastids, such as chloroplasts, for expression. Thus, the nucleic acid sequence, or sequences, encoding a gymnosperm monoterpene synthase protein, or proteins, may be inserted into the plastid for expression with appropriate plastid constructs and regulatory elements. Alternatively, nuclear transformation may be used in which case the expression cassette will contain a nucleic acid sequence encoding a transit peptide to direct the monoterpene biosynthesis enzyme of interest to the plastid. Such transit

peptides are known in the art. See, for example, Von Heijne et al. (1991) *Plant Mol. Biol. Rep.* **9**:104-126; Clark et al. (1989) *J. Biol. Chem.* **264**:17544-17550; della-Cioppa et al. (1987) *Plant Physiol.* **84**:965-968; Romer et al. (1993) *Biochem. Biophys. Res. Commun.* **196**:1414-1421; and, Shah et al. (1986) *Science* **233**:478-481. Nucleic acid sequences encoding gymnosperm monoterpene synthases of the present invention may utilize native or heterologous transit peptides.

The construct may also include any other necessary regulators such as plant translational consensus sequences (Joshi, C.P., (1987), *Nucleic Acids Research*, **15**:6643-6653), introns (Luehrsen and Walbot, (1991), *Mol. Gen. Genet.*, **225**:81-93) and the like, operably linked to a nucleotide sequence encoding a monoterpene synthase of the present invention.

It may be beneficial to include 5' leader sequences in the expression cassette which can act to enhance translation. Translation leaders are known in the art and include: picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5' noncoding region) (Elroy-Stein, O., Fuerst, T.R., and Moss, B. (1989) *PNAS USA* **86**:6126-6130); potyvirus leaders, for example, TEV leader (Tobacco Etch Virus) (Allison et al. (1986); MDMV leader (Maize Dwarf Mosaic Virus); *Virology*, **154**:9-20), and human immunoglobulin heavy-chain binding protein (BiP), (Macejak, D.G., and Sarnow, P. (1991), *Nature*, **353**:90-94; untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4), (Jobling, S.A., and Gehrke, L. (1987), *Nature*, **325**:622-625; tobacco mosaic virus leader (TMV), (Gallie, D.R. et al. (1989), *Molecular Biology of RNA*, pages 237-256; and maize chlorotic mottle virus leader (MCMV) (Lommel, S.A. et al. (1991), *Virology*, **81**:382-385. See also, Della-Cioppa et al., (1987), *Plant Physiology*, **84**:965-968.

Depending upon where the monoterpene synthase sequence of interest is to be expressed, it may be desirable to synthesize the sequence with plant preferred codons, or alternatively with chloroplast preferred codons. The plant preferred codons may be determined from the codons of highest frequency in the proteins expressed in the largest amount in the particular plant species of interest. See, EPA 0359472; EPA 0385962; WO 91/16432; Perlak et al. (1991) *Proc. Natl. Acad. Sci. USA* **88**:3324-3328; and Murray et al. (1989) *Nucleic Acids Research* **17**:477-498. In this manner, the nucleotide sequences can be optimized for expression in any plant. It is recognized that all or any part of the nucleic acid sequence encoding a gymnosperm monoterpene synthase protein may be optimized or synthetic. That is,

synthetic or partially optimized sequences may also be used. For the construction of chloroplast preferred genes, see U.S. Patent No. 5,545,817.

In preparing the transcription cassette, the various DNA fragments may be manipulated, so as to provide for the DNA sequences in the proper orientation and in the proper reading frame. Towards this end, adapters or linkers may be employed to join the DNA fragments or other manipulations may be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this purpose, *in vitro* mutagenesis, primer repair, restriction, annealing, resection, ligation, or the like may be employed, where insertions, deletions or substitutions, such as transitions and transversions, may be involved.

The recombinant DNA molecules of the invention can be introduced into the plant cell in a number of art-recognized ways. Those skilled in the art will appreciate that the choice of method might depend on the type of plant, *i.e.*, monocot or dicot, targeted for transformation. Suitable methods of transforming plant cells include microinjection (Crossway et al. (1986) *BioTechniques* **4**:320-334), electroporation (Riggs et al. (1986) *Proc. Natl. Acad. Sci. USA* **83**:5602-5606), *Agrobacterium* mediated transformation (Hinchee et al. (1988) *Biotechnology* **6**:915-921) and ballistic particle acceleration (see, for example, Sanford et al., U.S. Patent No. 4,945,050; and McCabe et al. (1988) *Biotechnology* **6**:923-926). Also see, Weissinger et al. (1988) *Annual Rev. Genet.* **22**:421-477; Sanford et al. (1987) *Particulate Science and Technology* **5**:27-37 (onion); Christou et al. (1988) *Plant Physiol.* **87**:671-674 (soybean); McCabe et al. (1988) *Bio/Technology* **6**:923-926 (soybean); Datta et al. (1990) *Biotechnology* **8**:736-740 (rice); Klein et al (1988) *Proc. Natl. Acad Sci. USA*, **85**:4305-4309 (maize); Klein et al. (1988) *Biotechnology* **6**:559-563 (maize); Klein et al. (1988) *Plant Physiol.* **91**:440-444 (maize); Fronun et al. (1990) *Biotechnology* **8**:833-839; and Gordon-Kamm et al. (1990) *Plant Cell* **2**:603-618 (maize).

Alternatively, a plant plastid can be transformed directly. Stable transformation of chloroplasts has been reported in higher plants, see, for example, SVAB et al. (1990) *Proc. Nat'l. Acad. Sci. USA* **87**:85268530; SVAB & Maliga (1993) *Proc. Natl. Acad. Sci. USA* **90**:913-917; Staub & Maliga (1993) *Embo J.* **12**:601-606. The method relies on particle gun delivery of DNA containing a selectable marker and targeting of the DNA to the plastid genome through homologous recombination. In such methods, plastid gene expression can be accomplished by use of a plastid gene promoter or by trans-activation of a silent

plastid-borne transgene positioned for expression from a selective promoter sequence such as that recognized by T7 RNA polymerase. The silent plastid gene is activated by expression of the specific RNA polymerase from a nuclear expression construct and targeting of the polymerase to the plastid by use of a transit peptide. Tissue-specific expression may be obtained in such a method by use of a nuclear-encoded and plastid-directed specific RNA polymerase expressed from a suitable plant tissue specific promoter. Such a system has been reported in McBride et al. (1994) *Proc. Natl. Acad Sci. USA* **91**:7301-7305.

The cells which have been transformed may be grown into plants by a variety of art-recognized means. See, for example, McConnick et al., *Plant Cell Reports* (1986), **5**:81-84. These plants may then be grown, and either selfed or crossed with a different plant strain, and the resulting homozygotes or hybrids having the desired phenotypic characteristic identified. Two or more generations may be grown to ensure that the subject phenotypic characteristic is stably maintained and inherited and then seeds harvested to ensure the desired phenotype or other property has been achieved.

As a host cell, any plant variety may be employed. Of particular interest, are plant species which provide seeds of commercial value. For the most part, plants will be chosen where the seed is produced in high amounts, a seed-specific product of interest is involved, or the seed or a seed part is edible. Seeds of interest in the practice of the present invention include, but are not limited to, the oil seeds, such as oilseed Brassica seeds, cotton seeds, soybean, safflower, sunflower, coconut, palm, and the like; grain seeds such as wheat, barley, oats, amaranth, flax, rye, triticale, rice and corn; other edible seeds or seeds with edible parts including pumpkin, squash, sesame, poppy, grape, mung beans, peanut, peas, beans, radish, alfalfa, cocoa, and coffee; and tree nuts such as walnuts, almonds, pecans, and chick-peas.

EXAMPLE 11

A Strategy For Cloning Gymnosperm Monoterpene Synthases

The present invention includes gymnosperm monoterpene synthase proteins, and nucleic acid molecules that encode gymnosperm monoterpene synthase proteins. The amino acid sequence of each of the gymnosperm monoterpene synthase proteins of the present invention each includes at least one of the amino acid sequence elements disclosed in Table 1.

Table 1

Amino Acids	Sequence	Orientation	Comments

1. 70-77	H S N (L, I, V) W D D D (SEQ ID NO: 46)	F only	HS makes a poor reverse primer
2. 148-153	A L D Y V Y (SEQ ID NO: 47)	F and R	
3. 306-312	E L A K L E F (SEQ ID NO: 48)	F and R	
4. 328-333	R W W K E S (SEQ ID NO: 49)	F and R	F primer uses 1st nt of Ser codon; R uses 1st two nts of Arg codon (rare only)
5. 377-383	(V, I, L) L D D M Y D (SEQ ID NO: 50)	F and R	
6. 377-383	(V, I, L) L D D L Y D (SEQ ID NO: 51)	F and R	Degeneracy of M/L at 381 too high for single primer
7. 377-383	(V, I, L) L D D I Y D (SEQ ID NO: 52)	F and R	
8. 543-549	C Y M K D (N, H) P (SEQ ID NO: 53)	R	F primer can also be constructed with this peptide but is too close to the 3' end to be useful

The numbers set forth in Table 1 for the first and last amino acid residue of each of the peptide sequences is the number of the corresponding amino acid residue in the amino acid sequence of the (-)-pinene synthase (SEQ ID NO:4) isolated from *Abies grandis*. Where a sequence of amino acid residues appears in brackets, *e.g.*, (L,I,V) in Table 1, the first amino acid residue within the brackets is the residue that appears in the (-)-pinene synthase amino acid sequence set forth in SEQ ID NO:4. The subsequent amino acid residues within the brackets represent other amino acid residues that commonly occur at the corresponding position in the amino acid sequence of other *Abies grandis* enzymes involved in terpene synthesis.

In Table 1, the letter "F" refers to the forward PCR reaction, *i.e.*, the PCR reaction which synthesizes the sense nucleic acid strand that encodes a gymnosperm monoterpene synthase. The letter "R" refers to the reverse PCR reaction, *i.e.*, the PCR reaction that synthesizes the antisense nucleic acid molecule that is complementary to the sense nucleic acid strand synthesized in the forward PCR reaction.

In order to clone nucleic acid molecules encoding gymnosperm monoterpene synthases of the present invention, one or more oligonucleotide molecules corresponding to at least a portion of one of the amino acid sequences set forth in Table 1 can be used as a probe or probes with which to screen a genomic or cDNA library derived from one or more gymnosperm species. In this context, the term "corresponding," or "correspond" or "corresponds," means that the oligonucleotide base sequence either a) encodes all or part of at least one of the amino acid sequences set forth in Table 1, or b) is complementary to a base sequence that encodes all or part of at least one of the amino acid sequences set forth in Table 1. The oligonucleotide probe(s) may contain a synthetic base, such as inosine, which can be substituted for one or more of the four, naturally-occurring bases, *i.e.*, adenine ("A"), guanine ("G"), cytosine ("C") and thymine ("T"). Thus, for example, the following oligonucleotide sequences "correspond" to the tripeptide sequence M M M: 5'ATGATGATG3' (sense orientation) (SEQ ID NO:54); 3'TACTACTAC5' (antisense orientation) (SEQ ID NO:55) and 3'IACIACIAC5' (SEQ ID NO:56).

One or more oligonucleotide sequence(s), corresponding to at least a portion of at least one of the amino acid sequences set forth in Table 1, can be used to screen a nucleic acid library in order to identify monoterpene synthase clones of the present invention, according to methods well known to one of ordinary skill in the art. *See, e.g., Sambrook et al, supra.* The stringency of the hybridization and wash conditions during library screening in accordance with the present invention, utilizing one or more oligonucleotide sequence(s) corresponding to at least a portion of at least one of the amino acid sequences set forth in Table 1, is at least: for the hybridization step, 6X SSPE, 40-45°C, for 36 hours; for the wash step, 3X SSPE, 45°C, 3 X 15 minute washes. The presently preferred hybridization and wash conditions during library screening, utilizing one or more oligonucleotide sequence(s) corresponding to at least a portion of at least one of the amino acid sequences set forth in Table 1, in accordance with the present invention are: for the hybridization step, 6X SSPE, 40-45°C, for 36 hours; for the wash step, 0.1X SSPE, 65°C-70°C, 3 X 15 minute washes.

Examples of oligonucleotide sequences, corresponding to at least one of the amino acid sequences set forth in Table 1, that hybridize, under the foregoing hybridization and wash conditions, to the sense strands of the nucleic acid sequences of the present invention that encode gymnosperm monoterpene synthase proteins are set forth in Table 2.

Table 2

Nucleic Acid Sequence	Corresponds to:
GTG TCG TTG GAG ACC CTG CTG CTG (SEQ ID NO:57)	SEQ ID No. 46
CGG GAG CTG ATG CAG ATG (SEQ ID NO:58)	SEQ ID No. 47
CTC GAG CGG TTC GAG CTC AAG (SEQ ID NO:59)	SEQ ID No. 48
GCC ACC ACC TTC CTC TCG (SEQ ID NO:60)	SEQ ID No. 49
GAG GAG CTG CTG TAC ATG CTG (SEQ ID NO:61)	SEQ ID No. 50
GAG GAG CTG CTG GAG ATG CTG (SEQ ID NO:62)	SEQ ID No. 51

Similarly, each of the gymnosperm monoterpene synthase clones set forth in SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:5, or a portion thereof, may be used as a probe to screen a nucleic acid library in order to isolate monoterpene synthase clones of the present invention, according to methods well known to one of ordinary skill in the art. *See, e.g., Sambrook et al, supra.* The stringency of the hybridization and wash conditions during library screening in accordance with the present invention is at least: for the hybridization step, 6X SSPE buffer at 45°C to 50°C for 36 hours; for the wash step, 3X SSPE buffer at 50°C (three, fifteen minute washes). In accordance with the present invention, the presently preferred hybridization and wash conditions during library screening utilizing any of the gymnosperm monoterpene synthase clones set forth in SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:5, or a portion thereof, as probe are: for the hybridization step, 6X SSPE, 40-45°C, for 36 hours; for the wash step, 0.1X SSPE, 70°C-75°C, 3 X 15 minute washes.

Additionally, at least two oligonucleotide sequence(s), each corresponding to at least a portion of at least one of the amino acid sequences set forth in Table 1, can be used in a PCR reaction to generate a portion of a monoterpene synthase clone of the present invention, which can be used as a probe to isolate a full-length clone of a monoterpene synthase clone of the present invention. Thus, oligonucleotides that are useful as probes in the forward PCR reaction correspond to at least a portion of at least one of the amino acid sequences disclosed in Table 1 as having the "F" orientation. Conversely, oligonucleotides that are useful as probes in the reverse

PCR reaction correspond to at least a portion of at least one of the amino acid sequences disclosed in Table 1 as having the "R" orientation. PCR reactions can be carried out according to art-recognized PCR reaction conditions, such as the PCR reaction conditions set forth in Example 1 herein and as set forth in "PCR Strategies", M.A. Innis, D.H. Gelfand and J.J. Sninsky, eds., 1995, Academic Press, San Diego, CA (Chapter 14); "PCR Protocols: A Guide to Methods and Applications", M.A. Innis, D.H. Gelfand, J.J. Sninsky and T.J. White, eds., Academic Press, NY (1990). The presently preferred PCR reaction conditions are:

dNTPs	200 μM each
$MgCl_2$	5-7 mM
F and R primers	100 nM - 1μM each
Taq polymerase	1-2 units/reaction
cDNA template	10-100 ng/reaction

Buffers, PCR grade water, and Chill-out wax or mineral oil

The presently preferred thermocycler conditions are:

Denaturation	94^0 x 2 min	1 cycle
Denaturation	94^0 x 45 s	35 cycles
Annealing	42^0 - 55^0 x 45 s - 1 min	"
Polymerization	72^0 x 1-2 min	"
Adenylation	72^0 x 10 min	1 cycle

While the preferred embodiment of the invention has been illustrated and described, it will be appreciated that various changes can be made therein without departing from the spirit and scope of the invention.

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. An isolated nucleotide sequence encoding a gymnosperm monoterpene synthase, said isolated nucleotide sequence being capable of hybridizing under stringent conditions to at least one oligonucleotide having a base sequence that corresponds to at least one peptide sequence selected from the group consisting of SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53.

2. An isolated nucleotide sequence of Claim 1 encoding a gymnosperm myrcene synthase.

3. An isolated nucleotide sequence of Claim 1 encoding a gymnosperm limonene synthase.

4. An isolated nucleotide sequence of Claim 1 encoding a gymnosperm pinene synthase.

5. An isolated nucleotide sequence of Claim 1 encoding a Grand fir (*Abies grandis*) monoterpene synthase.

6. An isolated nucleotide sequence of Claim 1 encoding a Grand fir (*Abies grandis*) myrcene synthase.

7. An isolated nucleotide sequence of Claim 1 encoding a Grand fir (*Abies grandis*) limonene synthase.

8. An isolated nucleotide sequence of Claim 1 encoding a Grand fir (*Abies grandis*) pinene synthase.

9. An isolated nucleotide sequence of Claim 1 having the sequence of SEQ ID NO:1.

10. An isolated nucleotide sequence of Claim 1 having the sequence of SEQ ID NO:3.

11. An isolated nucleotide sequence of Claim 1 having the sequence of SEQ ID NO:5

12. An isolated nucleotide sequence of Claim 1 which encodes the amino acid sequence of SEQ ID NO:2.

13. An isolated nucleotide sequence of Claim 1 which encodes the amino acid sequence of SEQ ID NO:4

14. An isolated nucleotide sequence of Claim 1 which encodes the amino acid sequence of SEQ ID NO:6.

15. An isolated nucleotide sequence encoding a gymnosperm monoterpene synthase protein, said isolated nucleotide sequence having a complementary nucleotide sequence that is capable of hybridizing to the nucleotide sequence of any one of SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:5.

16. An isolated nucleic acid molecule encoding a gymnosperm monoterpene synthase, said nucleic acid molecule being capable of hybridizing to at least two oligonucleotides, each of said oligonucleotides having a base sequence that corresponds to at least one peptide sequence selected from the group consisting of SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53.

17. An isolated nucleic acid molecule encoding a Grand fir (*Abies grandis*) monoterpene synthase, said nucleic acid molecule being capable of hybridizing to at least two oligonucleotides, each of said oligonucleotides having a base sequence that corresponds to at least one peptide sequence selected from the group consisting of SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53.

18. An isolated gymnosperm monoterpene synthase protein, said gymnosperm monoterpene synthase protein including at least one amino acid sequence element selected from the group consisting of SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53.

19. An isolated gymnosperm myrcene synthase protein, said gymnosperm myrcene synthase protein including at least one amino acid sequence element selected from the group consisting of SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53.

20. An isolated gymnosperm pinene synthase protein, said gymnosperm pinene synthase protein including at least one amino acid sequence element selected from the group consisting of SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53.

21. An isolated gymnosperm limonene synthase protein, said gymnosperm limonene synthase protein including at least one amino acid sequence element selected from the group consisting of SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53.

22. An isolated Grand fir (*Abies grandis*) monoterpene synthase protein, said Grand fir (*Abies grandis*) monoterpene synthase protein including at least one amino acid sequence element selected from the group consisting of SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53.

23. An isolated Grand fir (*Abies grandis*) monoterpene synthase protein of Claim 22, said monoterpene synthase protein having the sequence of SEQ ID NO:2.

24. An isolated Grand fir (*Abies grandis*) monoterpene synthase protein of Claim 22, said monoterpene synthase protein having the sequence of SEQ ID NO:4.

25. An isolated Grand fir (*Abies grandis*) monoterpene synthase protein of Claim 22, said monoterpene synthase protein having the sequence of SEQ ID NO:6.

26. A replicable expression vector comprising a nucleotide sequence encoding a gymnosperm monoterpene synthase, said nucleotide sequence encoding a gymnosperm monoterpene synthase being capable of hybridizing at high stringency to at least one oligonucleotide having a base sequence that corresponds to at least one peptide sequence selected from the group consisting of SEQ ID NO:46, SEQ ID

NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53.

27. A replicable expression vector of Claim 26 wherein said nucleotide sequence encoding a gymnosperm monoterpene synthase encodes a Grand fir (*Abies grandis*) monoterpene synthase.

28. A host cell comprising a vector of any one of claims 26 and 27.

29. A replicable expression vector comprising a nucleotide sequence encoding a gymnosperm monoterpene synthase, said nucleotide sequence having a complementary nucleotide sequence that is capable of hybridizing to the nucleotide sequence of any one of SEQ ID NO:1, SEQ ID NO:3 and SEQ ID NO:5.

30. A host cell comprising a vector of Claim 29.

31. A method of enhancing the production of a gymnosperm monoterpene synthase in a suitable host cell comprising introducing into the host cell an expression vector of Claim 26 under conditions enabling expression of the gymnosperm monoterpene synthase in the host cell.

32. A method for enhancing monoterpene levels in a host cell, said method comprising transforming said host cell with a nucleic acid construct comprising at least one nucleic acid sequence defining a transcriptional initiation region from a gene expressed in the host cell and at least one nucleic acid sequence encoding a gymnosperm monoterpene synthase, said nucleic acid sequence encoding a gymnosperm monoterpene synthase being capable of hybridizing at high stringency to at least one oligonucleotide having a base sequence that corresponds to at least one peptide sequence selected from the group consisting of SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53.

33. The method of Claim 32 wherein said nucleic acid sequence encoding a gymnosperm monoterpene synthase encodes a Grand fir (*Abies grandis*) monoterpene synthase.

34. A method for altering monoterpene levels in a seed from a host plant, said method comprising transforming said host plant with a nucleic acid construct

comprising at least one nucleic acid sequence defining a transcriptional initiation region from a gene expressed in a plant seed and at least one nucleic acid sequence encoding a gymnosperm monoterpene synthase, said nucleic acid sequence encoding a gymnosperm monoterpene synthase being capable of hybridizing at high stringency to at least one oligonucleotide having a base sequence that corresponds to at least one peptide sequence selected from the group consisting of SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53.

35. The method of Claim 34 wherein the transcriptional initiation region is from a gene preferentially expressed in a plant seed.

36. The method of Claim 34 wherein said nucleic acid sequence encoding a gymnosperm monoterpene synthase encodes a Grand fir (*Abies grandis*) monoterpene synthase.

37. The method of Claim 34 wherein said gymnosperm monoterpene synthase is not native to said host plant.

38. A transgenic plant produced according to the method of Claim 34.

39. A method for increasing the monoterpene biosynthetic flux in seed from a host plant, said method comprising transforming said host plant with a construct comprising as operably linked components, a transcriptional initiation region from a gene preferentially expressed in a plant seed, a plastid transit peptide, a DNA sequence encoding a gymnosperm monoterpene synthase protein, and a transcriptional termination region, said DNA sequence encoding a gymnosperm monoterpene synthase protein being capable of hybridizing at high stringency to at least one oligonucleotide having a base sequence that corresponds to at least one peptide sequence selected from the group consisting of SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53.

40. The method of Claim 39, wherein said DNA sequence encoding a gymnosperm monoterpene synthase protein encodes an enzyme selected from the group consisting of limonene synthase, myrcene synthase and pinene synthase.

41. A method for producing a monoterpene compound in a seed, said method comprising the steps of obtaining a transformed plant which produces said seed, said plant having and expressing in its genome:

a nucleic acid sequence encoding a gymnosperm monoterpene synthase protein which is operably linked to a plastid transit peptide and a transcriptional initiation region from a gene preferentially expressed in a plant seed, wherein said nucleic acid sequence encoding a gymnosperm monoterpene synthase protein is capable of hybridizing at high stringency to at least one oligonucleotide having a base sequence that corresponds to at least one peptide sequence selected from the group consisting of SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53.

42. A seed transformed with a nucleic acid sequence encoding a gymnosperm monoterpene synthase protein, said nucleic acid sequence encoding a gymnosperm monoterpene synthase protein being capable of hybridizing at high stringency to at least one oligonucleotide having a base sequence that corresponds to at least one peptide sequence selected from the group consisting of SEQ ID NO:SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52 and SEQ ID NO:53, said transformed seed having altered monoterpene levels.

43. The transformed seed of claim 42, wherein said seed produces increased levels of at least one monoterpene compound, said monoterpene compound selected from the group consisting of (-)-4S-limonene, (-)-α-pinene, myrcene, 4S-β-phellandrene.

44. The transformed seed of claim 43, wherein said seed produces increased levels of (-)4S-limonene.

45. Seed produced by the method of any one of claims 34, 35, 36, 37, 39, 40 and 41.

46. Plants produced by the method of any one of claims 34, 35, 36, 37, 39, 40 and 41.

47. Oil extracted from seeds produced by the method of any one of claims 34, 35, 36, 37, 39, 40 and 41.

48. Meal extracted from seed produced by the method of any one of claims 34, 35, 36, 37, 39, 40 and 41.

49. The method of any one of claims 34, 35, 36, 37, 39, 40 and 41, wherein said seed is from a plant selected from the group consisting of oilseed Brassica, cotton, soybean, safflower, sunflower, coconut, palm, wheat, barley, rice, corn, oats, amaranth, pumpkin, squash, sesame, poppy, grape, mung beans, peanut, peas, beans, radish, alfalfa, cocoa, coffee, and tree nuts.

50. The method of Claim 49, wherein said seed is from an oilseed crop plant selected from the group consisting of oilseed brassica, cotton, soybean, safflower, sunflower, palm, coconut, and corn.

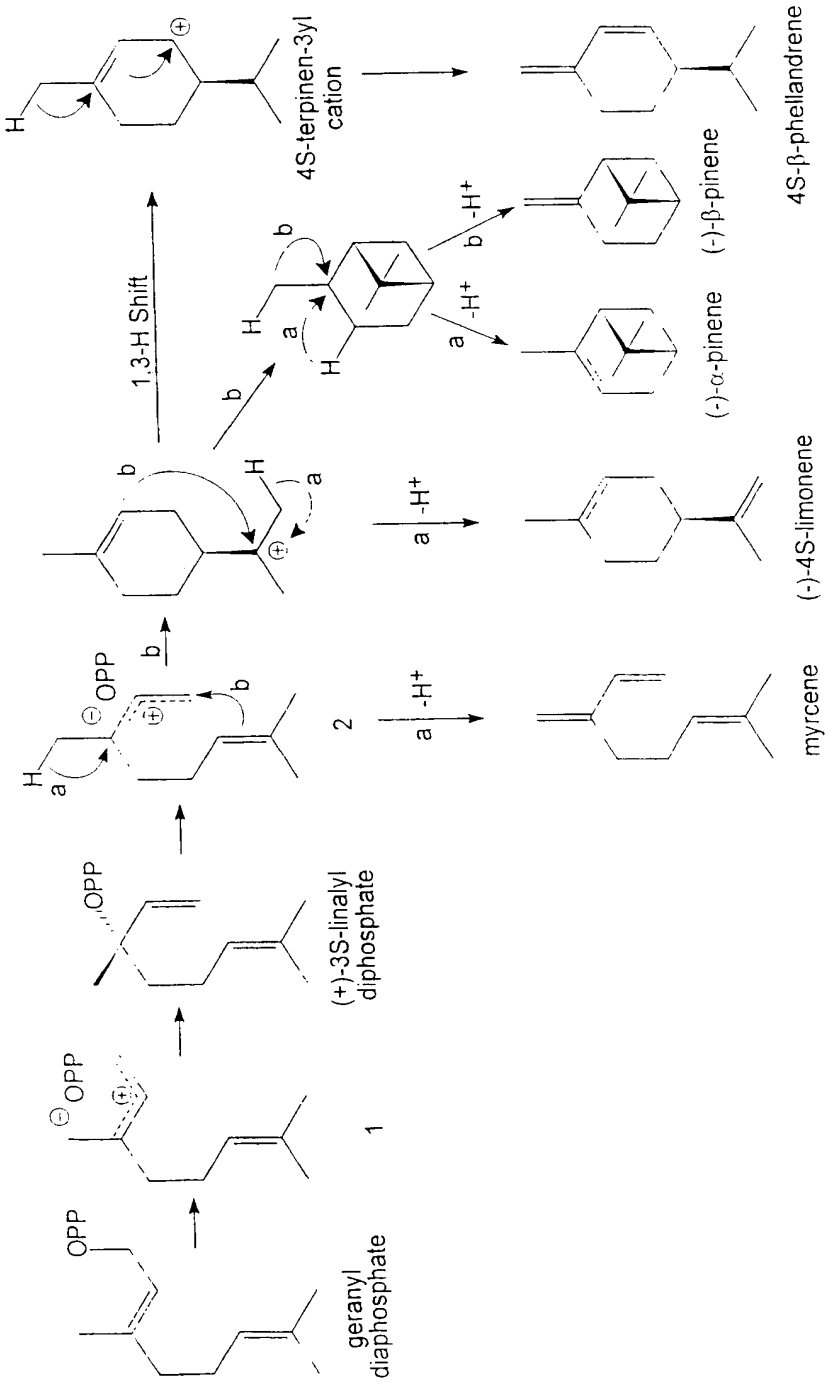
geranyl diphosphate
1
(+)-3S-linalyl diphosphate
2
myrcene
(-)-4S-limonene
1,3-H Shift
4S-terpinen-3yl cation
(-)-α-pinene
(-)-β-pinene
4S-β-phellandrene

Fig. 1

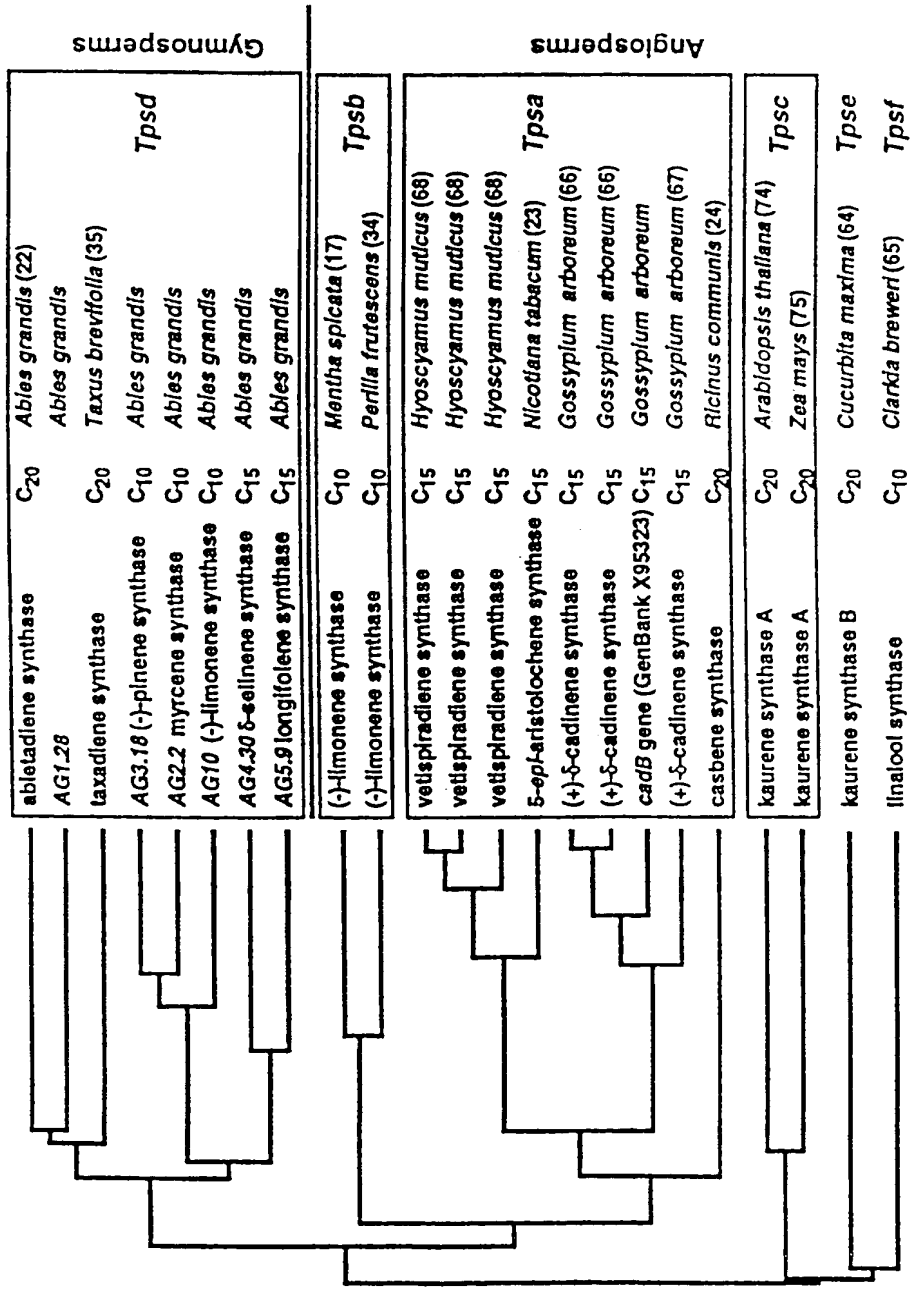
Gymnosperms
Angiosperms
abietadiene synthase C20 Abies grandis (22)
AG1.28 Abies grandis
taxadiene synthase C20 Taxus brevifolia (35)
AG3.18 (-)-pinene synthase C10 Abies grandis Tpsd
AG2.2 myrcene synthase C10 Abies grandis
AG10 (-)-limonene synthase C10 Abies grandis
AG4.30 δ-selinene synthase C15 Abies grandis
AG5.9 longifolene synthase C15 Abies grandis
(-)-limonene synthase C10 Mentha spicata (17) Tpsb
(-)-limonene synthase C10 Perilla frutescens (34)
vetispiradiene synthase C15 Hyoscyamus muticus (68)
vetispiradiene synthase C15 Hyoscyamus muticus (68)
vetispiradiene synthase C15 Hyoscyamus muticus (68)
5-epi-aristolochene synthase C15 Nicotiana tabacum (23) Tpsa
(+)-δ-cadinene synthase C15 Gossypium arboreum (66)
(+)-δ-cadinene synthase C15 Gossypium arboreum (66)
cadB gene (GenBank X95323) C15 Gossypium arboreum
(+)-δ-cadinene synthase C15 Gossypium arboreum (67)
casbene synthase C20 Ricinus communis (24)
kaurene synthase A C20 Arabidopsis thaliana (74) Tpsc
kaurene synthase A C20 Zea mays (75)
kaurene synthase B C20 Cucurbita maxima (64) Tpse
linalool synthase C10 Clarkia brewerl (65) Tpsf
40%
100%
amino acid identity

Fig. 2

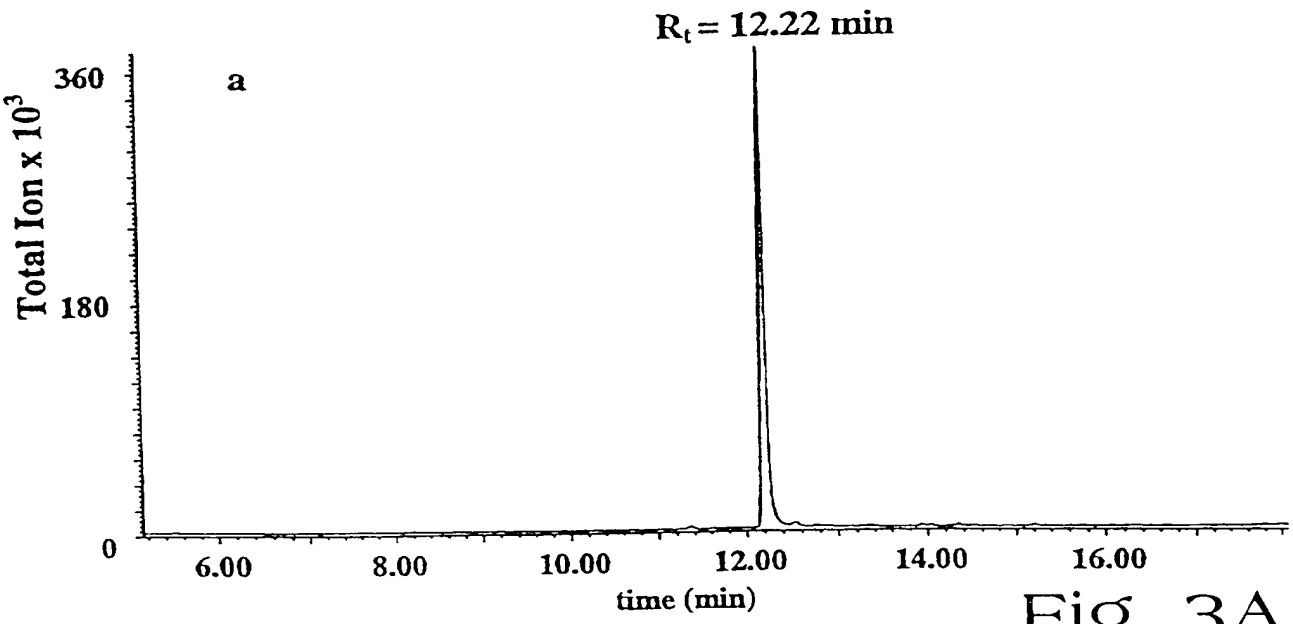
R_t = 12.22 min
a
360
180
0
Total Ion x 10^3
6.00
8.00
10.00
12.00
14.00
16.00
time (min)

Fig. 3A

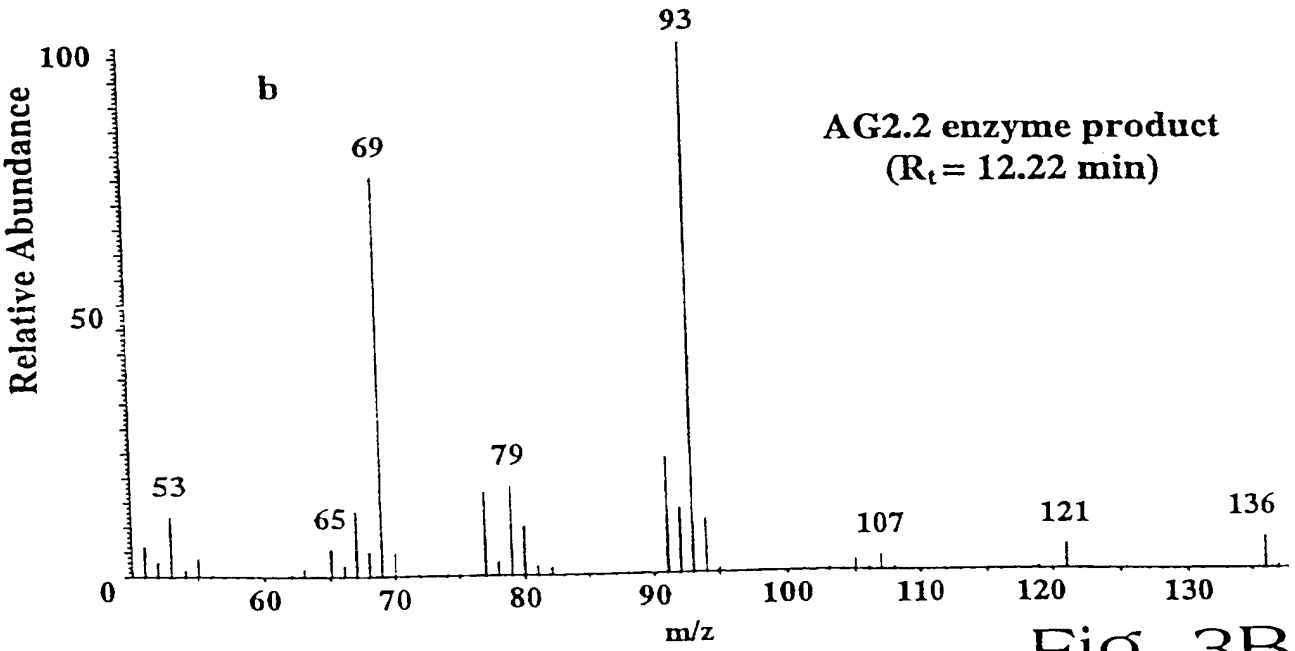
b
100
50
0
Relative Abundance
93
69
79
53
65
107
121
136
60
70
80
90
100
110
120
130
m/z
AG2.2 enzyme product
(R_t = 12.22 min)

Fig. 3B

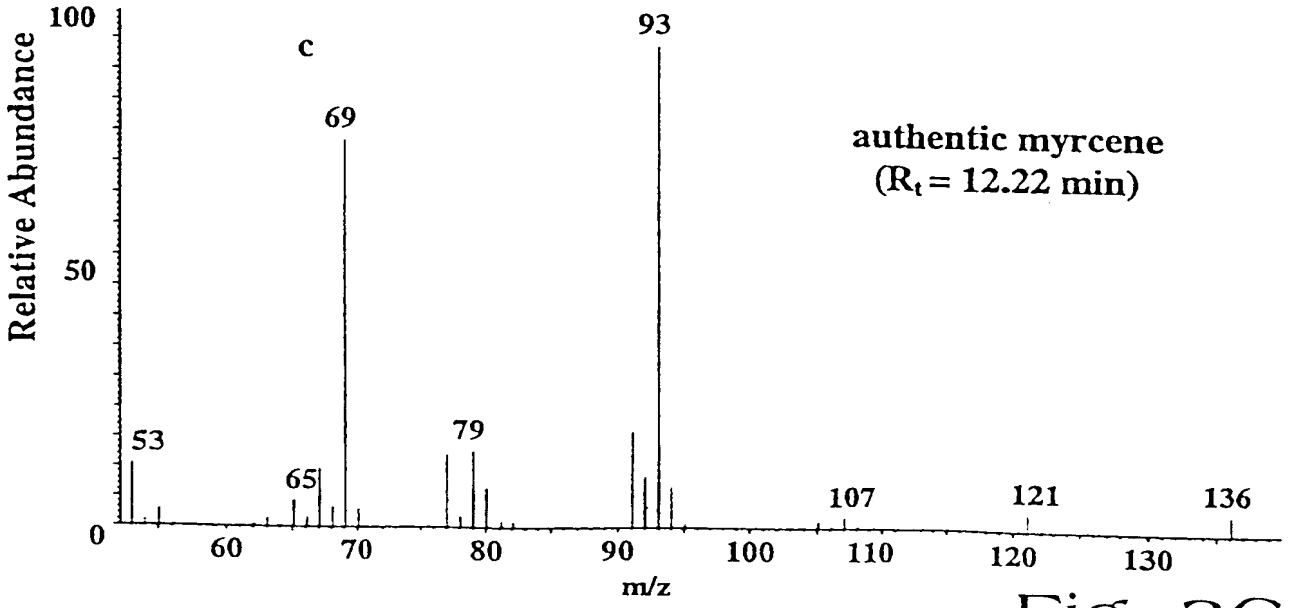

Fig. 3C

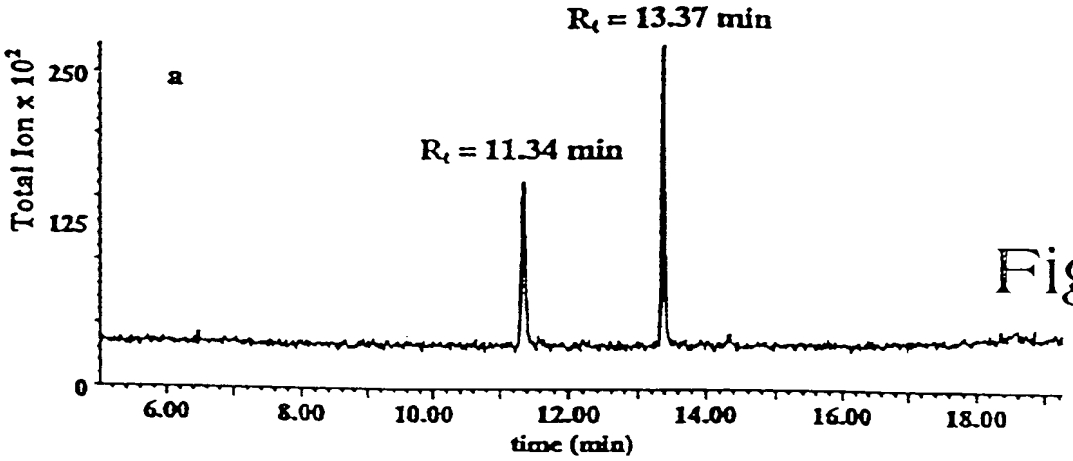

Fig. 4A

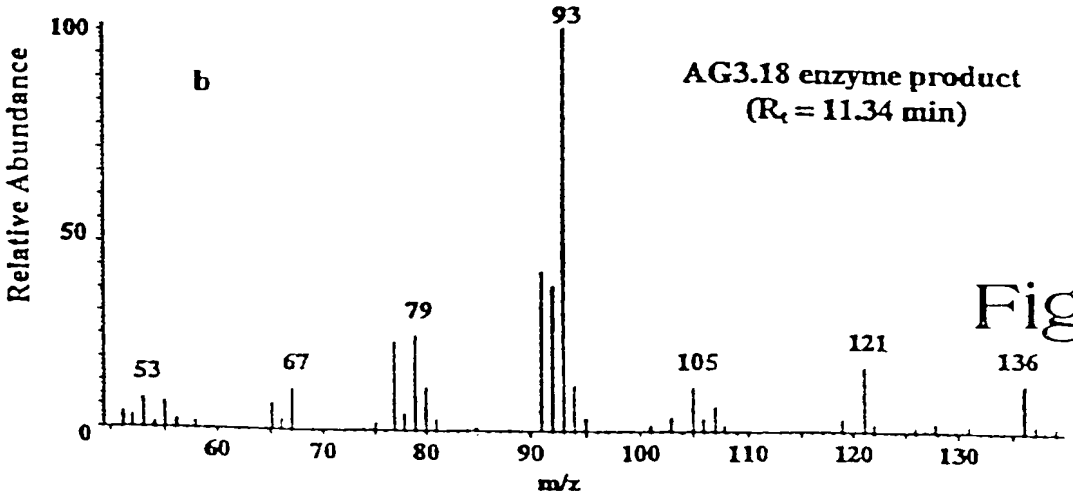

Fig. 4B

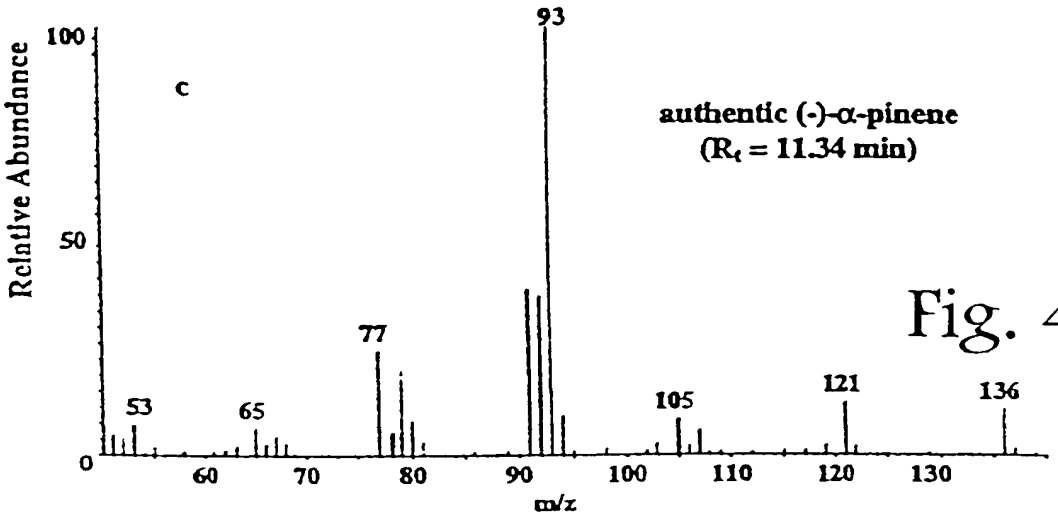

Fig. 4C

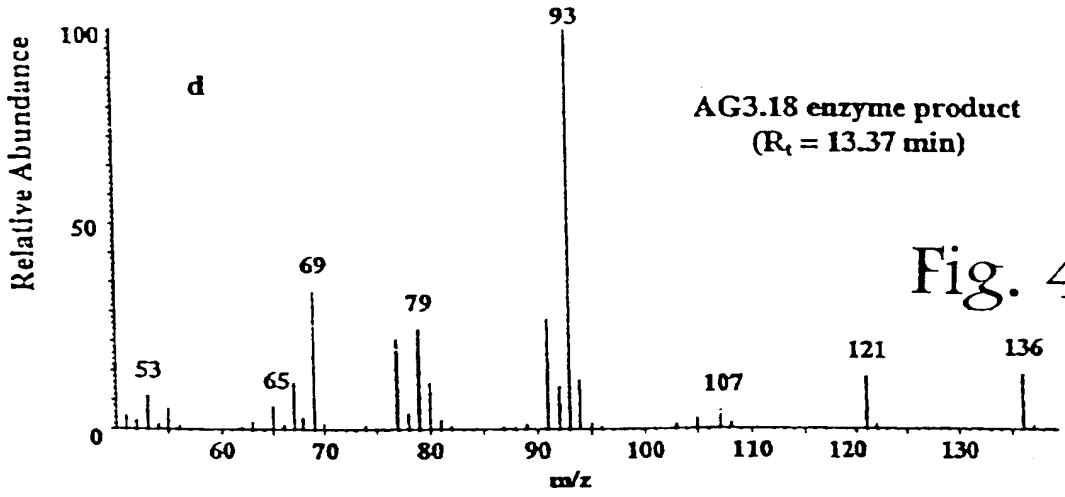

Fig. 4D

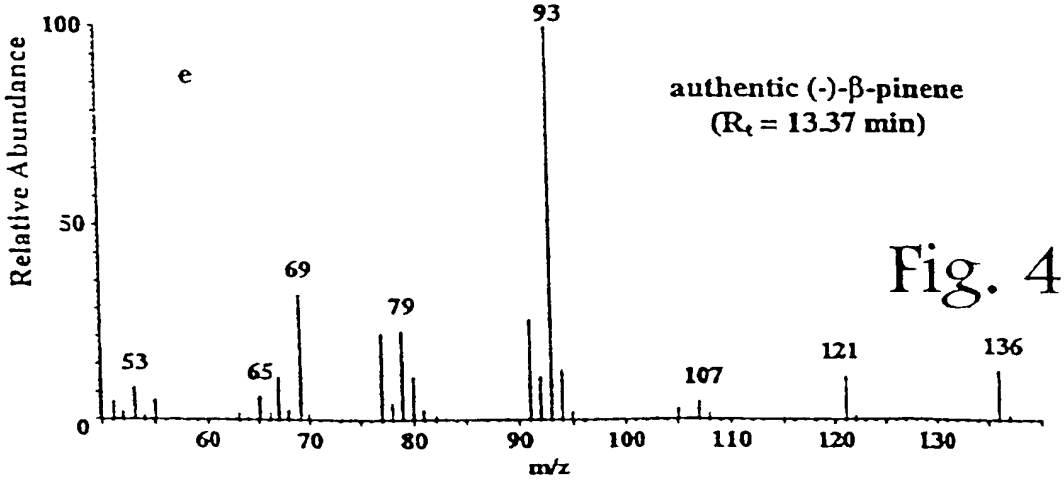

Fig. 4E

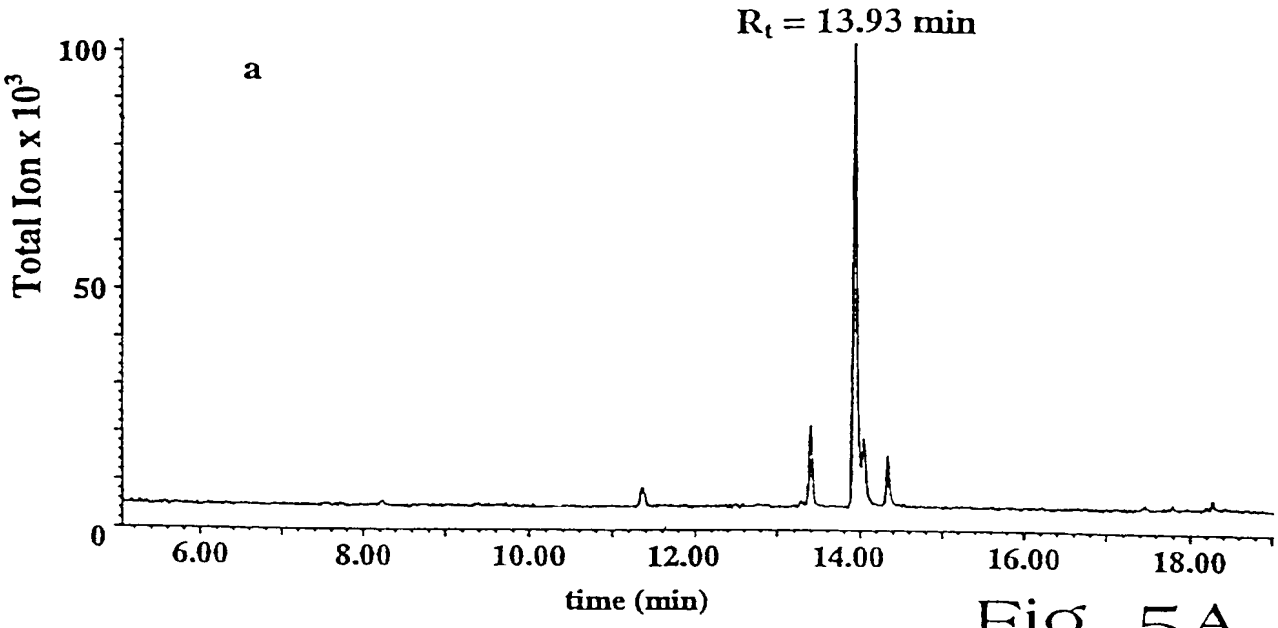
R_t = 13.93 min
a
100
50
0
Total Ion x 10^3
6.00
8.00
10.00
12.00
14.00
16.00
18.00
time (min)

Fig. 5A

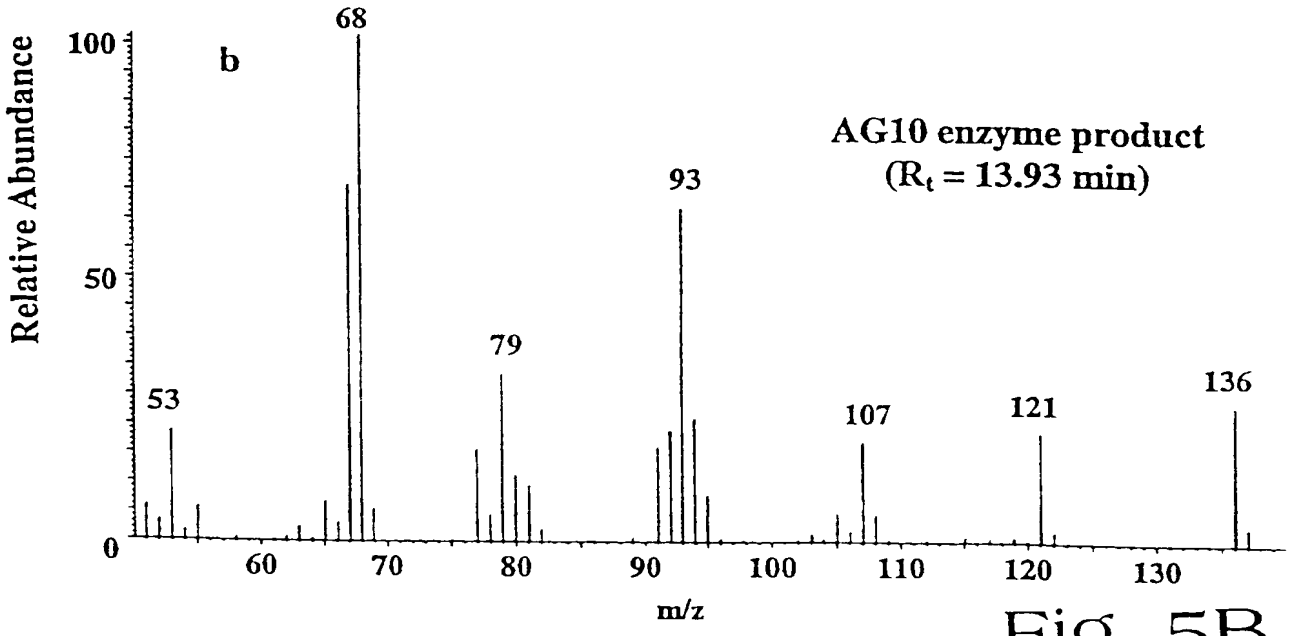
b
100
50
0
Relative Abundance
68
93
79
53
107
121
136
AG10 enzyme product
(R_t = 13.93 min)
60
70
80
90
100
110
120
130
m/z

Fig. 5B

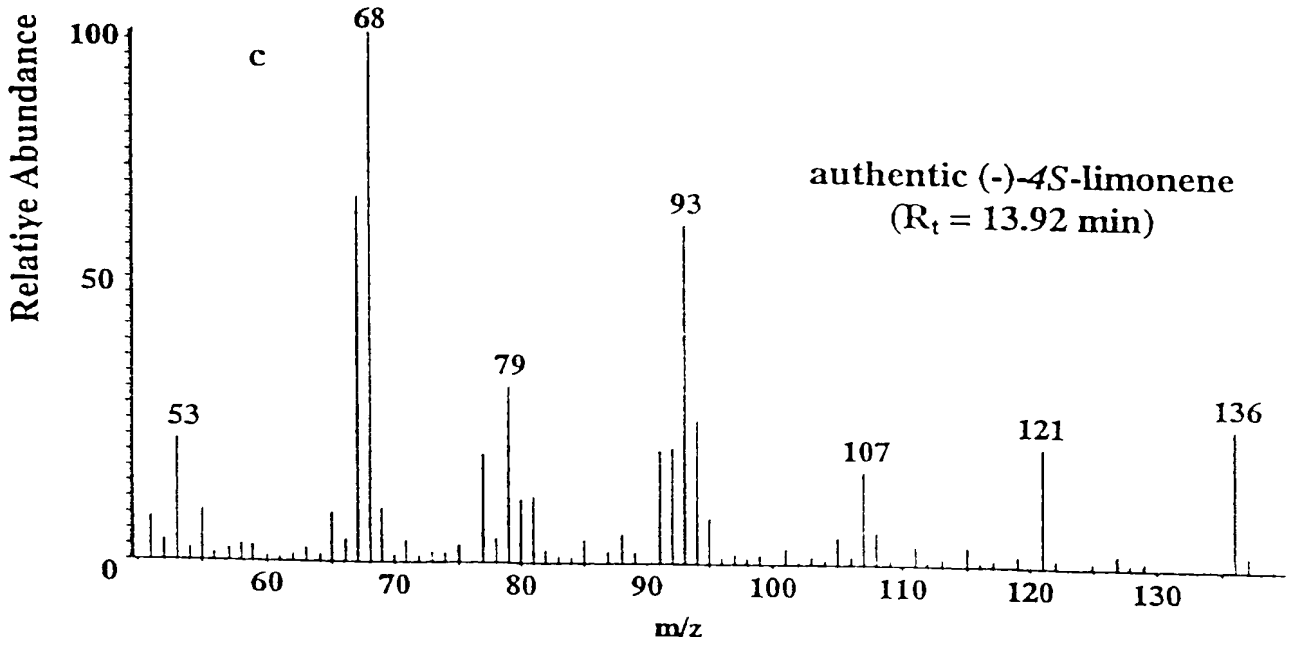
Relative Abundance
100
50
0
c
68
93
79
53
107
121
136
60
70
80
90
100
110
120
130
m/z
authentic (-)-4S-limonene
(R_t = 13.92 min)

Fig. 5C

SEQUENCE LISTING

<110> Croteau, Rodney B
Bohlmann, Carl J
Steele, Christopher L

<120> Monoterpene Synthases from Grand Fir (Abies grandis)

<130> WSUR112537

<140>
<141>

<150> 60/052,249
<151> 1997-07-11

<160> 62

<170> PatentIn Ver. 2.0

<210> 1
<211> 2196
<212> DNA
<213> Abies grandis

<220>
<221> CDS
<222> (69)..(1952)
<223> Clone AG2.2 encoding myrcene synthase

<400> 1
tgccggcacg aggttatctt gagcttcctc catataggcc aacacatatc atatcaaagg 60

gagcaaga atg gct ctg gtt tct atc tca ccg ttg gct tcg aaa tct tgc 110
Met Ala Leu Val Ser Ile Ser Pro Leu Ala Ser Lys Ser Cys
1 5 10

ctg cgc aag tcg ttg atc agt tca att cat gaa cat aag cct ccc tat 158
Leu Arg Lys Ser Leu Ile Ser Ser Ile His Glu His Lys Pro Pro Tyr
15 20 25 30

aga aca atc cca aat ctt gga atg cgt agg cga ggg aaa tct gtc acg 206
Arg Thr Ile Pro Asn Leu Gly Met Arg Arg Arg Gly Lys Ser Val Thr
35 40 45

cct tcc atg agc atc agt ttg gcc acc gct gca cct gat gat ggt gta 254
Pro Ser Met Ser Ile Ser Leu Ala Thr Ala Ala Pro Asp Asp Gly Val
50 55 60

caa aga cgc ata ggt gac tac cat tcc aat atc tgg gac gat gat ttc 302
Gln Arg Arg Ile Gly Asp Tyr His Ser Asn Ile Trp Asp Asp Asp Phe
65 70 75

ata cag tct cta tca acg cct tat ggg gaa ccc tct tac cag gaa cgt 350
Ile Gln Ser Leu Ser Thr Pro Tyr Gly Glu Pro Ser Tyr Gln Glu Arg
80 85 90

gct gag aga tta att gtg gag gta aag aag ata ttc aat tca atg tac 398
Ala Glu Arg Leu Ile Val Glu Val Lys Lys Ile Phe Asn Ser Met Tyr
95 100 105 110

```
ctg gat gat gga aga tta atg agt tcc ttt aat gat ctc atg caa cgc   446
Leu Asp Asp Gly Arg Leu Met Ser Ser Phe Asn Asp Leu Met Gln Arg
                115                 120                 125

ctt tgg ata gtc gat agc gtt gaa cgt ttg ggg ata gct aga cat ttc   494
Leu Trp Ile Val Asp Ser Val Glu Arg Leu Gly Ile Ala Arg His Phe
            130                 135                 140

aag aac gag ata aca tca gct ctg gat tat gtt ttc cgt tac tgg gag   542
Lys Asn Glu Ile Thr Ser Ala Leu Asp Tyr Val Phe Arg Tyr Trp Glu
        145                 150                 155

gaa aac ggc att gga tgt ggg aga gac agt att gtt act gat ctc aac   590
Glu Asn Gly Ile Gly Cys Gly Arg Asp Ser Ile Val Thr Asp Leu Asn
    160                 165                 170

tca act gcg ttg ggg ttt cga act ctt cga tta cac ggg tac act gta   638
Ser Thr Ala Leu Gly Phe Arg Thr Leu Arg Leu His Gly Tyr Thr Val
175                 180                 185                 190

tct cca gag gtt tta aaa gct ttt caa gat caa aat gga cag ttt gta   686
Ser Pro Glu Val Leu Lys Ala Phe Gln Asp Gln Asn Gly Gln Phe Val
                195                 200                 205

tgc tcc ccc ggt cag aca gag ggt gag atc aga agc gtt ctt aac tta   734
Cys Ser Pro Gly Gln Thr Glu Gly Glu Ile Arg Ser Val Leu Asn Leu
            210                 215                 220

tat cgg gct tcc ctc att gcc ttc cct ggt gag aaa gtt atg gaa gaa   782
Tyr Arg Ala Ser Leu Ile Ala Phe Pro Gly Glu Lys Val Met Glu Glu
        225                 230                 235

gct gaa atc ttc tcc aca aga tat ttg aaa gaa gct cta caa aag att   830
Ala Glu Ile Phe Ser Thr Arg Tyr Leu Lys Glu Ala Leu Gln Lys Ile
    240                 245                 250

cca gtc tcc gct ctt tca caa gag ata aag ttt gtt atg gaa tat ggc   878
Pro Val Ser Ala Leu Ser Gln Glu Ile Lys Phe Val Met Glu Tyr Gly
255                 260                 265                 270

tgg cac aca aat ttg cca aga ttg gaa gca aga aat tac ata gac aca   926
Trp His Thr Asn Leu Pro Arg Leu Glu Ala Arg Asn Tyr Ile Asp Thr
                275                 280                 285

ctt gag aaa gac acc agt gca tgg ctc aat aaa aat gct ggg aag aag   974
Leu Glu Lys Asp Thr Ser Ala Trp Leu Asn Lys Asn Ala Gly Lys Lys
            290                 295                 300

ctt tta gaa ctt gca aaa ttg gag ttc aat ata ttt aac tcc tta caa   1022
Leu Leu Glu Leu Ala Lys Leu Glu Phe Asn Ile Phe Asn Ser Leu Gln
        305                 310                 315

caa aag gaa tta caa tat ctt ttg aga tgg tgg aaa gag tcg gat ttg   1070
Gln Lys Glu Leu Gln Tyr Leu Leu Arg Trp Trp Lys Glu Ser Asp Leu
    320                 325                 330

cct aaa ttg aca ttt gct cgg cat cgt cat gtg gaa ttc tac act ttg   1118
Pro Lys Leu Thr Phe Ala Arg His Arg His Val Glu Phe Tyr Thr Leu
335                 340                 345                 350
```

```
gcc tct tgt att gcc att gac cca aaa cat tct gca ttc aga cta ggc   1166
Ala Ser Cys Ile Ala Ile Asp Pro Lys His Ser Ala Phe Arg Leu Gly
                355                 360                 365

ttc gcc aaa atg tgt cat ctt gtc aca gtt ttg gac gat att tac gac   1214
Phe Ala Lys Met Cys His Leu Val Thr Val Leu Asp Asp Ile Tyr Asp
            370                 375                 380

act ttt gga acg att gac gag ctt gaa ctc ttc aca tct gca att aag   1262
Thr Phe Gly Thr Ile Asp Glu Leu Glu Leu Phe Thr Ser Ala Ile Lys
        385                 390                 395

aga tgg aat tca tca gag ata gaa cac ctt cca gaa tat atg aaa tgt   1310
Arg Trp Asn Ser Ser Glu Ile Glu His Leu Pro Glu Tyr Met Lys Cys
    400                 405                 410

gtg tac atg gtc gtg ttt gaa act gta aat gaa ctg aca cga gag gcg   1358
Val Tyr Met Val Val Phe Glu Thr Val Asn Glu Leu Thr Arg Glu Ala
415                 420                 425                 430

gag aag act caa ggg aga aac act ctc aac tat gtt cga aag gct tgg   1406
Glu Lys Thr Gln Gly Arg Asn Thr Leu Asn Tyr Val Arg Lys Ala Trp
                435                 440                 445

gag gct tat ttt gat tca tat atg gaa gaa gca aaa tgg atc tct aat   1454
Glu Ala Tyr Phe Asp Ser Tyr Met Glu Glu Ala Lys Trp Ile Ser Asn
            450                 455                 460

ggt tat ctg cca atg ttt gaa gag tac cat gag aat ggg aaa gtg agc   1502
Gly Tyr Leu Pro Met Phe Glu Glu Tyr His Glu Asn Gly Lys Val Ser
        465                 470                 475

tct gca tat cgc gta gca aca ttg caa ccc atc ctc act ttg aat gca   1550
Ser Ala Tyr Arg Val Ala Thr Leu Gln Pro Ile Leu Thr Leu Asn Ala
    480                 485                 490

tgg ctt cct gat tac atc ttg aag gga att gat ttt cca tcc agg ttc   1598
Trp Leu Pro Asp Tyr Ile Leu Lys Gly Ile Asp Phe Pro Ser Arg Phe
495                 500                 505                 510

aat gat ttg gca tcg tcc ttc ctt cgg cta cga ggt gac aca cgc tgc   1646
Asn Asp Leu Ala Ser Ser Phe Leu Arg Leu Arg Gly Asp Thr Arg Cys
                515                 520                 525

tac aag gcc gat agg gat cgt ggt gaa gaa gct tcg tgt ata tca tgt   1694
Tyr Lys Ala Asp Arg Asp Arg Gly Glu Glu Ala Ser Cys Ile Ser Cys
            530                 535                 540

tat atg aaa gac aat cct gga tca acc gaa gaa gat gcc ctc aat cat   1742
Tyr Met Lys Asp Asn Pro Gly Ser Thr Glu Glu Asp Ala Leu Asn His
        545                 550                 555

atc aat gcc atg gtc aat gac ata atc aaa gaa tta aat tgg gaa ctt   1790
Ile Asn Ala Met Val Asn Asp Ile Ile Lys Glu Leu Asn Trp Glu Leu
    560                 565                 570

cta aga tcc aac gac aat att cca atg ctg gcc aag aaa cat gct ttt   1838
Leu Arg Ser Asn Asp Asn Ile Pro Met Leu Ala Lys Lys His Ala Phe
575                 580                 585                 590
```

```
gac ata aca aga gct ctc cac cat ctc tac ata tat cga gat ggc ttt   1886
Asp Ile Thr Arg Ala Leu His His Leu Tyr Ile Tyr Arg Asp Gly Phe
                595                 600                 605

agt gtt gcc aac aag gaa aca aaa aaa ttg gtt atg gaa aca ctc ctt   1934
Ser Val Ala Asn Lys Glu Thr Lys Lys Leu Val Met Glu Thr Leu Leu
            610                 615                 620

gaa tct atg ctt ttt taa ctataaccat atccataata ataagctcat          1982
Glu Ser Met Leu Phe
        625

aatgctaaat tattggcctt atgacatagt ttatgtatgt acttgtgtga attcaatcat 2042

atcgtgtggg tatgattaaa aagctagagc ttactaggtt agtaacatgg tgataaaagt 2102

tataaaatgt gagttataga gatacccatg ttgaataatg aattacaaaa agagaaattt 2162

atgtagaata agattggaag cttttcaatt gttt                             2196
```

<210> 2
<211> 627
<212> PRT
<213> Abies grandis

<400> 2

```
Met Ala Leu Val Ser Ile Ser Pro Leu Ala Ser Lys Ser Cys Leu Arg
  1               5                  10                  15

Lys Ser Leu Ile Ser Ser Ile His Glu His Lys Pro Pro Tyr Arg Thr
             20                  25                  30

Ile Pro Asn Leu Gly Met Arg Arg Arg Gly Lys Ser Val Thr Pro Ser
         35                  40                  45

Met Ser Ile Ser Leu Ala Thr Ala Ala Pro Asp Asp Gly Val Gln Arg
     50                  55                  60

Arg Ile Gly Asp Tyr His Ser Asn Ile Trp Asp Asp Asp Phe Ile Gln
 65                  70                  75                  80

Ser Leu Ser Thr Pro Tyr Gly Glu Pro Ser Tyr Gln Glu Arg Ala Glu
                 85                  90                  95

Arg Leu Ile Val Glu Val Lys Lys Ile Phe Asn Ser Met Tyr Leu Asp
            100                 105                 110

Asp Gly Arg Leu Met Ser Ser Phe Asn Asp Leu Met Gln Arg Leu Trp
        115                 120                 125

Ile Val Asp Ser Val Glu Arg Leu Gly Ile Ala Arg His Phe Lys Asn
    130                 135                 140

Glu Ile Thr Ser Ala Leu Asp Tyr Val Phe Arg Tyr Trp Glu Glu Asn
145                 150                 155                 160

Gly Ile Gly Cys Gly Arg Asp Ser Ile Val Thr Asp Leu Asn Ser Thr
                165                 170                 175
```

Ala Leu Gly Phe Arg Thr Leu Arg Leu His Gly Tyr Thr Val Ser Pro
180 185 190

Glu Val Leu Lys Ala Phe Gln Asp Gln Asn Gly Gln Phe Val Cys Ser
195 200 205

Pro Gly Gln Thr Glu Gly Glu Ile Arg Ser Val Leu Asn Leu Tyr Arg
210 215 220

Ala Ser Leu Ile Ala Phe Pro Gly Glu Lys Val Met Glu Glu Ala Glu
225 230 235 240

Ile Phe Ser Thr Arg Tyr Leu Lys Glu Ala Leu Gln Lys Ile Pro Val
245 250 255

Ser Ala Leu Ser Gln Glu Ile Lys Phe Val Met Glu Tyr Gly Trp His
260 265 270

Thr Asn Leu Pro Arg Leu Glu Ala Arg Asn Tyr Ile Asp Thr Leu Glu
275 280 285

Lys Asp Thr Ser Ala Trp Leu Asn Lys Asn Ala Gly Lys Lys Leu Leu
290 295 300

Glu Leu Ala Lys Leu Glu Phe Asn Ile Phe Asn Ser Leu Gln Gln Lys
305 310 315 320

Glu Leu Gln Tyr Leu Leu Arg Trp Trp Lys Glu Ser Asp Leu Pro Lys
325 330 335

Leu Thr Phe Ala Arg His Arg His Val Glu Phe Tyr Thr Leu Ala Ser
340 345 350

Cys Ile Ala Ile Asp Pro Lys His Ser Ala Phe Arg Leu Gly Phe Ala
355 360 365

Lys Met Cys His Leu Val Thr Val Leu Asp Asp Ile Tyr Asp Thr Phe
370 375 380

Gly Thr Ile Asp Glu Leu Glu Leu Phe Thr Ser Ala Ile Lys Arg Trp
385 390 395 400

Asn Ser Ser Glu Ile Glu His Leu Pro Glu Tyr Met Lys Cys Val Tyr
405 410 415

Met Val Val Phe Glu Thr Val Asn Glu Leu Thr Arg Glu Ala Glu Lys
420 425 430

Thr Gln Gly Arg Asn Thr Leu Asn Tyr Val Arg Lys Ala Trp Glu Ala
435 440 445

Tyr Phe Asp Ser Tyr Met Glu Glu Ala Lys Trp Ile Ser Asn Gly Tyr
450 455 460

Leu Pro Met Phe Glu Glu Tyr His Glu Asn Gly Lys Val Ser Ser Ala
465 470 475 480

Tyr Arg Val Ala Thr Leu Gln Pro Ile Leu Thr Leu Asn Ala Trp Leu
485 490 495

```
Pro Asp Tyr Ile Leu Lys Gly Ile Asp Phe Pro Ser Arg Phe Asn Asp
            500                 505                 510

Leu Ala Ser Ser Phe Leu Arg Leu Arg Gly Asp Thr Arg Cys Tyr Lys
        515                 520                 525

Ala Asp Arg Asp Arg Gly Glu Glu Ala Ser Cys Ile Ser Cys Tyr Met
    530                 535                 540

Lys Asp Asn Pro Gly Ser Thr Glu Glu Asp Ala Leu Asn His Ile Asn
545                 550                 555                 560

Ala Met Val Asn Asp Ile Ile Lys Glu Leu Asn Trp Glu Leu Leu Arg
                565                 570                 575

Ser Asn Asp Asn Ile Pro Met Leu Ala Lys Lys His Ala Phe Asp Ile
            580                 585                 590

Thr Arg Ala Leu His His Leu Tyr Ile Tyr Arg Asp Gly Phe Ser Val
        595                 600                 605

Ala Asn Lys Glu Thr Lys Lys Leu Val Met Glu Thr Leu Leu Glu Ser
    610                 615                 620

Met Leu Phe
625
```

<210> 3
<211> 2018
<212> DNA
<213> Abies grandis

<220>
<221> CDS
<222> (6)..(1892)
<223> Clone AG3.18 encoding pinene synthase

<400> 3

```
cagca atg gct cta gtt tct acc gca ccg ttg gct tcc aaa tca tgc ctg 50
      Met Ala Leu Val Ser Thr Ala Pro Leu Ala Ser Lys Ser Cys Leu
        1               5                  10                  15

cac aaa tcg ttg atc agt tct acc cat gag ctt aag gct ctc tct aga   98
His Lys Ser Leu Ile Ser Ser Thr His Glu Leu Lys Ala Leu Ser Arg
                20                  25                  30

aca att cca gct cta gga atg agt agg cga ggg aaa tct atc act cct   146
Thr Ile Pro Ala Leu Gly Met Ser Arg Arg Gly Lys Ser Ile Thr Pro
            35                  40                  45

tcc atc agc atg agc tct acc acc gtt gta acc gat gat ggt gta cga   194
Ser Ile Ser Met Ser Ser Thr Thr Val Val Thr Asp Asp Gly Val Arg
        50                  55                  60

aga cgc atg ggc gat ttc cat tcc aac ctc tgg gac gat gat gtc ata   242
Arg Arg Met Gly Asp Phe His Ser Asn Leu Trp Asp Asp Asp Val Ile
    65                  70                  75
```

```
cag tct tta cca acg gct tat gag gaa aaa tcg tac ctg gag cgt gct   290
Gln Ser Leu Pro Thr Ala Tyr Glu Glu Lys Ser Tyr Leu Glu Arg Ala
 80                  85                  90                  95

gag aaa ctg atc ggg gaa gta aag aac atg ttc aat tcg atg tca tta   338
Glu Lys Leu Ile Gly Glu Val Lys Asn Met Phe Asn Ser Met Ser Leu
                100                 105                 110

gaa gat gga gag tta atg agt ccg ctc aat gat ctc att caa cgc ctt   386
Glu Asp Gly Glu Leu Met Ser Pro Leu Asn Asp Leu Ile Gln Arg Leu
            115                 120                 125

tgg att gtc gac agc ctt gaa cgt ttg ggg atc cat aga cat ttc aaa   434
Trp Ile Val Asp Ser Leu Glu Arg Leu Gly Ile His Arg His Phe Lys
        130                 135                 140

gat gag ata aaa tcg gcg ctt gat tat gtt tac agt tat tgg ggc gaa   482
Asp Glu Ile Lys Ser Ala Leu Asp Tyr Val Tyr Ser Tyr Trp Gly Glu
    145                 150                 155

aat ggc atc gga tgc ggg agg gag agt gtt gtt act gat ctg aac tca   530
Asn Gly Ile Gly Cys Gly Arg Glu Ser Val Val Thr Asp Leu Asn Ser
160                 165                 170                 175

act gcg ttg ggg ctt cga acc cta cga cta cac gga tac ccg gtg tct   578
Thr Ala Leu Gly Leu Arg Thr Leu Arg Leu His Gly Tyr Pro Val Ser
                180                 185                 190

tca gat gtt ttc aaa gct ttc aaa ggc caa aat ggg cag ttt tcc tgc   626
Ser Asp Val Phe Lys Ala Phe Lys Gly Gln Asn Gly Gln Phe Ser Cys
            195                 200                 205

tct gaa aat att cag aca gat gaa gag atc aga ggc gtt ctg aat tta   674
Ser Glu Asn Ile Gln Thr Asp Glu Glu Ile Arg Gly Val Leu Asn Leu
        210                 215                 220

ttc cgg gcc tcc ctc att gcc ttt cca ggg gag aaa att atg gat gag   722
Phe Arg Ala Ser Leu Ile Ala Phe Pro Gly Glu Lys Ile Met Asp Glu
    225                 230                 235

gct gaa atc ttc tct acc aaa tat tta aaa gaa gcc ctg caa aag att   770
Ala Glu Ile Phe Ser Thr Lys Tyr Leu Lys Glu Ala Leu Gln Lys Ile
240                 245                 250                 255

ccg gtc tcc agt ctt tcg cga gag atc ggg gac gtt ttg gaa tat ggt   818
Pro Val Ser Ser Leu Ser Arg Glu Ile Gly Asp Val Leu Glu Tyr Gly
                260                 265                 270

tgg cac aca tat ttg ccg cga ttg gaa gca agg aat tac atc caa gtc   866
Trp His Thr Tyr Leu Pro Arg Leu Glu Ala Arg Asn Tyr Ile Gln Val
            275                 280                 285

ttt gga cag gac act gag aac acg aag tca tat gtg aag agc aaa aaa   914
Phe Gly Gln Asp Thr Glu Asn Thr Lys Ser Tyr Val Lys Ser Lys Lys
        290                 295                 300

ctt tta gaa ctc gca aaa ttg gag ttc aac atc ttt caa tcc tta caa   962
Leu Leu Glu Leu Ala Lys Leu Glu Phe Asn Ile Phe Gln Ser Leu Gln
    305                 310                 315
```

```
aag agg gag tta gaa agt ctg gtc aga tgg tgg aaa gaa tcg ggt ttt    1010
Lys Arg Glu Leu Glu Ser Leu Val Arg Trp Trp Lys Glu Ser Gly Phe
320                 325                 330                 335

cct gag atg acc ttc tgc cga cat cgt cac gtg gaa tac tac act ttg    1058
Pro Glu Met Thr Phe Cys Arg His Arg His Val Glu Tyr Tyr Thr Leu
                340                 345                 350

gct tcc tgc att gcg ttc gag cct caa cat tct gga ttc aga ctc ggc    1106
Ala Ser Cys Ile Ala Phe Glu Pro Gln His Ser Gly Phe Arg Leu Gly
            355                 360                 365

ttt gcc aag acg tgt cat ctt atc acg gtt ctt gac gat atg tac gac    1154
Phe Ala Lys Thr Cys His Leu Ile Thr Val Leu Asp Asp Met Tyr Asp
        370                 375                 380

acc ttc ggc aca gta gac gag ctg gaa ctc ttc aca gcg aca atg aag    1202
Thr Phe Gly Thr Val Asp Glu Leu Glu Leu Phe Thr Ala Thr Met Lys
    385                 390                 395

aga tgg gat ccg tcc tcg ata gat tgc ctt cca gaa tat atg aaa gga    1250
Arg Trp Asp Pro Ser Ser Ile Asp Cys Leu Pro Glu Tyr Met Lys Gly
400                 405                 410                 415

gtg tac ata gcg gtt tac gac acc gta aat gaa atg gct cga gag gca    1298
Val Tyr Ile Ala Val Tyr Asp Thr Val Asn Glu Met Ala Arg Glu Ala
                420                 425                 430

gag gag gct caa ggc cga gat acg ctc aca tat gct cgg gaa gct tgg    1346
Glu Glu Ala Gln Gly Arg Asp Thr Leu Thr Tyr Ala Arg Glu Ala Trp
            435                 440                 445

gag gct tat att gat tcg tat atg caa gaa gca agg tgg atc gcc act    1394
Glu Ala Tyr Ile Asp Ser Tyr Met Gln Glu Ala Arg Trp Ile Ala Thr
        450                 455                 460

ggt tac ctg ccc tcc ttt gat gag tac tac gag aat ggg aaa gtt agc    1442
Gly Tyr Leu Pro Ser Phe Asp Glu Tyr Tyr Glu Asn Gly Lys Val Ser
    465                 470                 475

tgt ggt cat cgc ata tcc gca ttg caa ccc att ctg aca atg gac atc    1490
Cys Gly His Arg Ile Ser Ala Leu Gln Pro Ile Leu Thr Met Asp Ile
480                 485                 490                 495

ccc ttt cct gat cat atc ctc aag gaa gtt gac ttc cca tca aag ctt    1538
Pro Phe Pro Asp His Ile Leu Lys Glu Val Asp Phe Pro Ser Lys Leu
                500                 505                 510

aac gac ttg gca tgt gcc atc ctt cga tta cga ggt gat acg cgg tgc    1586
Asn Asp Leu Ala Cys Ala Ile Leu Arg Leu Arg Gly Asp Thr Arg Cys
            515                 520                 525

tac aag gcg gac agg gct cgt gga gaa gaa gct tcc tct ata tca tgt    1634
Tyr Lys Ala Asp Arg Ala Arg Gly Glu Glu Ala Ser Ser Ile Ser Cys
        530                 535                 540

tat atg aaa gac aat cct gga gta tca gag gaa gat gct ctc gat cat    1682
Tyr Met Lys Asp Asn Pro Gly Val Ser Glu Glu Asp Ala Leu Asp His
    545                 550                 555
```

```
atc aac gcc atg atc agt gac gta atc aaa gga tta aat tgg gaa ctt   1730
Ile Asn Ala Met Ile Ser Asp Val Ile Lys Gly Leu Asn Trp Glu Leu
560                 565                 570                 575

ctc aaa cca gac atc aat gtt ccc atc tcg gcg aag aaa cat gct ttt   1778
Leu Lys Pro Asp Ile Asn Val Pro Ile Ser Ala Lys Lys His Ala Phe
                580                 585                 590

gac atc gcc aga gct ttc cat tac ggc tac aaa tac cga gac ggc tac   1826
Asp Ile Ala Arg Ala Phe His Tyr Gly Tyr Lys Tyr Arg Asp Gly Tyr
            595                 600                 605

agc gtt gcc aac gtt gaa acg aag agt ttg gtc acg aga acc ctc ctt   1874
Ser Val Ala Asn Val Glu Thr Lys Ser Leu Val Thr Arg Thr Leu Leu
        610                 615                 620

gaa tct gtg cct ttg tag caacagctca aatctatgcc ctatgctatg          1922
Glu Ser Val Pro Leu
    625

tcgggttaaa atatatgtgg aaggtagccg ttggatgtag aggataagtt tgttataatt 1982

taataaagtt gtaatttaaa aaaaaaaaaa aaaaaa                           2018
```

<210> 4
<211> 628
<212> PRT
<213> Abies grandis

<400> 4

```
Met Ala Leu Val Ser Thr Ala Pro Leu Ala Ser Lys Ser Cys Leu His
  1               5                  10                  15

Lys Ser Leu Ile Ser Ser Thr His Glu Leu Lys Ala Leu Ser Arg Thr
             20                  25                  30

Ile Pro Ala Leu Gly Met Ser Arg Arg Gly Lys Ser Ile Thr Pro Ser
         35                  40                  45

Ile Ser Met Ser Ser Thr Thr Val Val Thr Asp Asp Gly Val Arg Arg
     50                  55                  60

Arg Met Gly Asp Phe His Ser Asn Leu Trp Asp Asp Asp Val Ile Gln
 65                  70                  75                  80

Ser Leu Pro Thr Ala Tyr Glu Glu Lys Ser Tyr Leu Glu Arg Ala Glu
                 85                  90                  95

Lys Leu Ile Gly Glu Val Lys Asn Met Phe Asn Ser Met Ser Leu Glu
            100                 105                 110

Asp Gly Glu Leu Met Ser Pro Leu Asn Asp Leu Ile Gln Arg Leu Trp
        115                 120                 125

Ile Val Asp Ser Leu Glu Arg Leu Gly Ile His Arg His Phe Lys Asp
    130                 135                 140

Glu Ile Lys Ser Ala Leu Asp Tyr Val Tyr Ser Tyr Trp Gly Glu Asn
145                 150                 155                 160
```

Gly Ile Gly Cys Gly Arg Glu Ser Val Val Thr Asp Leu Asn Ser Thr
165 170 175

Ala Leu Gly Leu Arg Thr Leu Arg Leu His Gly Tyr Pro Val Ser Ser
180 185 190

Asp Val Phe Lys Ala Phe Lys Gly Gln Asn Gly Gln Phe Ser Cys Ser
195 200 205

Glu Asn Ile Gln Thr Asp Glu Glu Ile Arg Gly Val Leu Asn Leu Phe
210 215 220

Arg Ala Ser Leu Ile Ala Phe Pro Gly Glu Lys Ile Met Asp Glu Ala
225 230 235 240

Glu Ile Phe Ser Thr Lys Tyr Leu Lys Glu Ala Leu Gln Lys Ile Pro
245 250 255

Val Ser Ser Leu Ser Arg Glu Ile Gly Asp Val Leu Glu Tyr Gly Trp
260 265 270

His Thr Tyr Leu Pro Arg Leu Glu Ala Arg Asn Tyr Ile Gln Val Phe
275 280 285

Gly Gln Asp Thr Glu Asn Thr Lys Ser Tyr Val Lys Ser Lys Lys Leu
290 295 300

Leu Glu Leu Ala Lys Leu Glu Phe Asn Ile Phe Gln Ser Leu Gln Lys
305 310 315 320

Arg Glu Leu Glu Ser Leu Val Arg Trp Trp Lys Glu Ser Gly Phe Pro
325 330 335

Glu Met Thr Phe Cys Arg His Arg His Val Glu Tyr Tyr Thr Leu Ala
340 345 350

Ser Cys Ile Ala Phe Glu Pro Gln His Ser Gly Phe Arg Leu Gly Phe
355 360 365

Ala Lys Thr Cys His Leu Ile Thr Val Leu Asp Asp Met Tyr Asp Thr
370 375 380

Phe Gly Thr Val Asp Glu Leu Glu Leu Phe Thr Ala Thr Met Lys Arg
385 390 395 400

Trp Asp Pro Ser Ser Ile Asp Cys Leu Pro Glu Tyr Met Lys Gly Val
405 410 415

Tyr Ile Ala Val Tyr Asp Thr Val Asn Glu Met Ala Arg Glu Ala Glu
420 425 430

Glu Ala Gln Gly Arg Asp Thr Leu Thr Tyr Ala Arg Glu Ala Trp Glu
435 440 445

Ala Tyr Ile Asp Ser Tyr Met Gln Glu Ala Arg Trp Ile Ala Thr Gly
450 455 460

Tyr Leu Pro Ser Phe Asp Glu Tyr Tyr Glu Asn Gly Lys Val Ser Cys
465 470 475 480

```
Gly His Arg Ile Ser Ala Leu Gln Pro Ile Leu Thr Met Asp Ile Pro
                485                 490                 495

Phe Pro Asp His Ile Leu Lys Glu Val Asp Phe Pro Ser Lys Leu Asn
            500                 505                 510

Asp Leu Ala Cys Ala Ile Leu Arg Leu Arg Gly Asp Thr Arg Cys Tyr
        515                 520                 525

Lys Ala Asp Arg Ala Arg Gly Glu Glu Ala Ser Ser Ile Ser Cys Tyr
    530                 535                 540

Met Lys Asp Asn Pro Gly Val Ser Glu Glu Asp Ala Leu Asp His Ile
545                 550                 555                 560

Asn Ala Met Ile Ser Asp Val Ile Lys Gly Leu Asn Trp Glu Leu Leu
                565                 570                 575

Lys Pro Asp Ile Asn Val Pro Ile Ser Ala Lys Lys His Ala Phe Asp
            580                 585                 590

Ile Ala Arg Ala Phe His Tyr Gly Tyr Lys Tyr Arg Asp Gly Tyr Ser
        595                 600                 605

Val Ala Asn Val Glu Thr Lys Ser Leu Val Thr Arg Thr Leu Leu Glu
    610                 615                 620

Ser Val Pro Leu
625
```

<210> 5
<211> 2089
<212> DNA
<213> Abies grandis

<220>
<221> CDS
<222> (73)..(1986)
<223> Clone AG10 encoding limonene synthase

<400> 5

```
tgccgtttaa tcggtttaaa gaagctacca tagttcggtt taaagaagct accatagttt 60

aggcaggaat cc atg gct ctc ctt tct atc gta tct ttg cag gtt ccc aaa 111
              Met Ala Leu Leu Ser Ile Val Ser Leu Gln Val Pro Lys
                1               5                  10

tcc tgc ggg ctg aaa tcg ttg atc agt tcc agc aat gtg cag aag gct   159
Ser Cys Gly Leu Lys Ser Leu Ile Ser Ser Ser Asn Val Gln Lys Ala
     15                  20                  25

ctc tgt atc tct aca gca gtc cca aca ctc aga atg cgt agg cga cag   207
Leu Cys Ile Ser Thr Ala Val Pro Thr Leu Arg Met Arg Arg Arg Gln
 30                  35                  40                  45

aaa gct ctg gtc atc aac atg aaa ttg acc act gta tcc cat cgt gat   255
Lys Ala Leu Val Ile Asn Met Lys Leu Thr Thr Val Ser His Arg Asp
                50                  55                  60
```

```
gat aat ggt ggt ggt gta ctg caa aga cgc ata gcc gat cat cat ccc   303
Asp Asn Gly Gly Gly Val Leu Gln Arg Arg Ile Ala Asp His His Pro
            65                  70                  75

aac ctg tgg gaa gat gat ttc ata caa tca ttg tcc tca cct tat ggg   351
Asn Leu Trp Glu Asp Asp Phe Ile Gln Ser Leu Ser Ser Pro Tyr Gly
        80                  85                  90

gga tct tcg tac agt gaa cgt gct gag aca gtc gtt gag gaa gta aaa   399
Gly Ser Ser Tyr Ser Glu Arg Ala Glu Thr Val Val Glu Glu Val Lys
    95                  100                 105

gag atg ttc aat tca ata cca aat aat aga gaa tta ttt ggt tcc caa   447
Glu Met Phe Asn Ser Ile Pro Asn Asn Arg Glu Leu Phe Gly Ser Gln
110                 115                 120                 125

aat gat ctc ctt aca cgc ctt tgg atg gtg gat agc att gaa cgt ctg   495
Asn Asp Leu Leu Thr Arg Leu Trp Met Val Asp Ser Ile Glu Arg Leu
                130                 135                 140

ggg ata gat aga cat ttc caa aat gag ata aga gta gcc ctc gat tat   543
Gly Ile Asp Arg His Phe Gln Asn Glu Ile Arg Val Ala Leu Asp Tyr
            145                 150                 155

gtt tac agt tat tgg aag gaa aag gaa ggc att ggg tgt ggc aga gat   591
Val Tyr Ser Tyr Trp Lys Glu Lys Glu Gly Ile Gly Cys Gly Arg Asp
        160                 165                 170

tct act ttt cct gat ctc aac tcg act gcc ttg gcg ctt cga act ctt   639
Ser Thr Phe Pro Asp Leu Asn Ser Thr Ala Leu Ala Leu Arg Thr Leu
    175                 180                 185

cga ctg cac gga tac aat gtg tct tca gat gtg ctg gaa tac ttc aaa   687
Arg Leu His Gly Tyr Asn Val Ser Ser Asp Val Leu Glu Tyr Phe Lys
190                 195                 200                 205

gat gaa aag ggg cat ttt gcc tgc cct gca atc cta acc gag gga cag   735
Asp Glu Lys Gly His Phe Ala Cys Pro Ala Ile Leu Thr Glu Gly Gln
                210                 215                 220

atc act aga agt gtt cta aat tta tat cgg gct tcc ctg gtc gcc ttt   783
Ile Thr Arg Ser Val Leu Asn Leu Tyr Arg Ala Ser Leu Val Ala Phe
            225                 230                 235

ccc ggg gag aaa gtt atg gaa gag gct gaa atc ttc tcg gca tct tat   831
Pro Gly Glu Lys Val Met Glu Glu Ala Glu Ile Phe Ser Ala Ser Tyr
        240                 245                 250

ttg aaa aaa gtc tta caa aag att ccg gtc tcc aat ctt tca gga gag   879
Leu Lys Lys Val Leu Gln Lys Ile Pro Val Ser Asn Leu Ser Gly Glu
    255                 260                 265

ata gaa tat gtt ttg gaa tat ggt tgg cac acg aat ttg ccg aga ttg   927
Ile Glu Tyr Val Leu Glu Tyr Gly Trp His Thr Asn Leu Pro Arg Leu
270                 275                 280                 285

gaa gca aga aat tat atc gag gtc tac gag cag agc ggc tat gaa agc   975
Glu Ala Arg Asn Tyr Ile Glu Val Tyr Glu Gln Ser Gly Tyr Glu Ser
                290                 295                 300
```

```
tta aac gag atg cca tat atg aac atg aag aag ctt tta caa ctt gca   1023
Leu Asn Glu Met Pro Tyr Met Asn Met Lys Lys Leu Leu Gln Leu Ala
            305                 310                 315

aaa ttg gag ttc aat atc ttt cac tct ttg caa cta aga gag tta caa   1071
Lys Leu Glu Phe Asn Ile Phe His Ser Leu Gln Leu Arg Glu Leu Gln
        320                 325                 330

tct atc tcc aga tgg tgg aaa gaa tca ggt tcg tct caa ctg act ttt   1119
Ser Ile Ser Arg Trp Trp Lys Glu Ser Gly Ser Ser Gln Leu Thr Phe
    335                 340                 345

aca cgg cat cgt cac gtg gaa tac tac act atg gca tct tgc att tct   1167
Thr Arg His Arg His Val Glu Tyr Tyr Thr Met Ala Ser Cys Ile Ser
350                 355                 360                 365

atg ttg cca aaa cat tca gct ttc aga atg gag ttt gtc aaa gtg tgt   1215
Met Leu Pro Lys His Ser Ala Phe Arg Met Glu Phe Val Lys Val Cys
                370                 375                 380

cat ctt gta aca gtt ctc gat gat ata tat gac act ttt gga aca atg   1263
His Leu Val Thr Val Leu Asp Asp Ile Tyr Asp Thr Phe Gly Thr Met
            385                 390                 395

aac gaa ctc caa ctt ttt acg gat gca att aag aga tgg gat ttg tca   1311
Asn Glu Leu Gln Leu Phe Thr Asp Ala Ile Lys Arg Trp Asp Leu Ser
        400                 405                 410

acg aca agg tgg ctt cca gaa tat atg aaa gga gtg tac atg gac ttg   1359
Thr Thr Arg Trp Leu Pro Glu Tyr Met Lys Gly Val Tyr Met Asp Leu
    415                 420                 425

tat caa tgc att aat gaa atg gtg gaa gag gct gag aag act caa ggc   1407
Tyr Gln Cys Ile Asn Glu Met Val Glu Glu Ala Glu Lys Thr Gln Gly
430                 435                 440                 445

cga gat atg ctc aac tat att caa aat gct tgg gaa gcc cta ttt gat   1455
Arg Asp Met Leu Asn Tyr Ile Gln Asn Ala Trp Glu Ala Leu Phe Asp
                450                 455                 460

acc ttt atg caa gaa gca aag tgg atc tcc agc agt tat ctc cca acg   1503
Thr Phe Met Gln Glu Ala Lys Trp Ile Ser Ser Ser Tyr Leu Pro Thr
            465                 470                 475

ttt gag gag tac ttg aag aat gca aaa gtt agt tct ggt tct cgc ata   1551
Phe Glu Glu Tyr Leu Lys Asn Ala Lys Val Ser Ser Gly Ser Arg Ile
        480                 485                 490

gcc aca tta caa ccc att ctc act ttg gat gta cca ctt cct gat tac   1599
Ala Thr Leu Gln Pro Ile Leu Thr Leu Asp Val Pro Leu Pro Asp Tyr
    495                 500                 505

ata ctg caa gaa att gat tat cca tcc aga ttc aat gag tta gct tcg   1647
Ile Leu Gln Glu Ile Asp Tyr Pro Ser Arg Phe Asn Glu Leu Ala Ser
510                 515                 520                 525

tcc atc ctt cga cta cga ggt gac acg cgc tgc tac aag gcg gat agg   1695
Ser Ile Leu Arg Leu Arg Gly Asp Thr Arg Cys Tyr Lys Ala Asp Arg
                530                 535                 540
```

```
gcc cgt gga gaa gaa gct tca gct ata tcg tgt tat atg aaa gac cat   1743
Ala Arg Gly Glu Glu Ala Ser Ala Ile Ser Cys Tyr Met Lys Asp His
            545                 550                 555

cct gga tca ata gag gaa gat gct ctc aat cat atc aac gcc atg atc   1791
Pro Gly Ser Ile Glu Glu Asp Ala Leu Asn His Ile Asn Ala Met Ile
        560                 565                 570

agt gat gca atc aga gaa tta aat tgg gag ctt ctc aga ccg gat agc   1839
Ser Asp Ala Ile Arg Glu Leu Asn Trp Glu Leu Leu Arg Pro Asp Ser
    575                 580                 585

aaa agt ccc atc tct tcc aag aaa cat gct ttt gac atc acc aga gct   1887
Lys Ser Pro Ile Ser Ser Lys Lys His Ala Phe Asp Ile Thr Arg Ala
590                 595                 600                 605

ttc cat cat gtc tac aaa tat cga gat ggt tac act gtt tcc aac aac   1935
Phe His His Val Tyr Lys Tyr Arg Asp Gly Tyr Thr Val Ser Asn Asn
                610                 615                 620

gaa aca aag aat ttg gtg atg aaa acc gtt ctt gaa cct ctc gct ttg   1983
Glu Thr Lys Asn Leu Val Met Lys Thr Val Leu Glu Pro Leu Ala Leu
            625                 630                 635

taa aaacatatag aatgcattaa aatgtgggaa gtctataatc tagactattc        2036

tctatctttc ataatgtaga tctggatgtg tattgaactc taaaaaaaaa aaa        2089
```

<210> 6
<211> 637
<212> PRT
<213> Abies grandis

<400> 6

```
Met Ala Leu Leu Ser Ile Val Ser Leu Gln Val Pro Lys Ser Cys Gly
  1               5                  10                  15

Leu Lys Ser Leu Ile Ser Ser Ser Asn Val Gln Lys Ala Leu Cys Ile
             20                  25                  30

Ser Thr Ala Val Pro Thr Leu Arg Met Arg Arg Arg Gln Lys Ala Leu
         35                  40                  45

Val Ile Asn Met Lys Leu Thr Thr Val Ser His Arg Asp Asp Asn Gly
     50                  55                  60

Gly Gly Val Leu Gln Arg Arg Ile Ala Asp His His Pro Asn Leu Trp
 65                  70                  75                  80

Glu Asp Asp Phe Ile Gln Ser Leu Ser Ser Pro Tyr Gly Gly Ser Ser
                 85                  90                  95

Tyr Ser Glu Arg Ala Glu Thr Val Val Glu Glu Val Lys Glu Met Phe
            100                 105                 110

Asn Ser Ile Pro Asn Asn Arg Glu Leu Phe Gly Ser Gln Asn Asp Leu
        115                 120                 125
```

```
Leu Thr Arg Leu Trp Met Val Asp Ser Ile Glu Arg Leu Gly Ile Asp
    130                 135                 140

Arg His Phe Gln Asn Glu Ile Arg Val Ala Leu Asp Tyr Val Tyr Ser
145                 150                 155                 160

Tyr Trp Lys Glu Lys Glu Gly Ile Gly Cys Gly Arg Asp Ser Thr Phe
                165                 170                 175

Pro Asp Leu Asn Ser Thr Ala Leu Ala Leu Arg Thr Leu Arg Leu His
            180                 185                 190

Gly Tyr Asn Val Ser Ser Asp Val Leu Glu Tyr Phe Lys Asp Glu Lys
        195                 200                 205

Gly His Phe Ala Cys Pro Ala Ile Leu Thr Glu Gly Gln Ile Thr Arg
    210                 215                 220

Ser Val Leu Asn Leu Tyr Arg Ala Ser Leu Val Ala Phe Pro Gly Glu
225                 230                 235                 240

Lys Val Met Glu Glu Ala Glu Ile Phe Ser Ala Ser Tyr Leu Lys Lys
                245                 250                 255

Val Leu Gln Lys Ile Pro Val Ser Asn Leu Ser Gly Glu Ile Glu Tyr
            260                 265                 270

Val Leu Glu Tyr Gly Trp His Thr Asn Leu Pro Arg Leu Glu Ala Arg
        275                 280                 285

Asn Tyr Ile Glu Val Tyr Glu Gln Ser Gly Tyr Glu Ser Leu Asn Glu
    290                 295                 300

Met Pro Tyr Met Asn Met Lys Lys Leu Leu Gln Leu Ala Lys Leu Glu
305                 310                 315                 320

Phe Asn Ile Phe His Ser Leu Gln Leu Arg Glu Leu Gln Ser Ile Ser
                325                 330                 335

Arg Trp Trp Lys Glu Ser Gly Ser Ser Gln Leu Thr Phe Thr Arg His
            340                 345                 350

Arg His Val Glu Tyr Tyr Thr Met Ala Ser Cys Ile Ser Met Leu Pro
        355                 360                 365

Lys His Ser Ala Phe Arg Met Glu Phe Val Lys Val Cys His Leu Val
    370                 375                 380

Thr Val Leu Asp Asp Ile Tyr Asp Thr Phe Gly Thr Met Asn Glu Leu
385                 390                 395                 400

Gln Leu Phe Thr Asp Ala Ile Lys Arg Trp Asp Leu Ser Thr Thr Arg
                405                 410                 415

Trp Leu Pro Glu Tyr Met Lys Gly Val Tyr Met Asp Leu Tyr Gln Cys
            420                 425                 430

Ile Asn Glu Met Val Glu Glu Ala Glu Lys Thr Gln Gly Arg Asp Met
        435                 440                 445
```

```
Leu Asn Tyr Ile Gln Asn Ala Trp Glu Ala Leu Phe Asp Thr Phe Met
     450                 455                 460

Gln Glu Ala Lys Trp Ile Ser Ser Ser Tyr Leu Pro Thr Phe Glu Glu
465                 470                 475                 480

Tyr Leu Lys Asn Ala Lys Val Ser Ser Gly Ser Arg Ile Ala Thr Leu
                485                 490                 495

Gln Pro Ile Leu Thr Leu Asp Val Pro Leu Pro Asp Tyr Ile Leu Gln
            500                 505                 510

Glu Ile Asp Tyr Pro Ser Arg Phe Asn Glu Leu Ala Ser Ser Ile Leu
        515                 520                 525

Arg Leu Arg Gly Asp Thr Arg Cys Tyr Lys Ala Asp Arg Ala Arg Gly
    530                 535                 540

Glu Glu Ala Ser Ala Ile Ser Cys Tyr Met Lys Asp His Pro Gly Ser
545                 550                 555                 560

Ile Glu Glu Asp Ala Leu Asn His Ile Asn Ala Met Ile Ser Asp Ala
                565                 570                 575

Ile Arg Glu Leu Asn Trp Glu Leu Leu Arg Pro Asp Ser Lys Ser Pro
            580                 585                 590

Ile Ser Ser Lys Lys His Ala Phe Asp Ile Thr Arg Ala Phe His His
        595                 600                 605

Val Tyr Lys Tyr Arg Asp Gly Tyr Thr Val Ser Asn Asn Glu Thr Lys
    610                 615                 620

Asn Leu Val Met Lys Thr Val Leu Glu Pro Leu Ala Leu
625                 630                 635
```

<210> 7
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Degenerate oligonucleotide PCR primer A wherein the letter "n" indicates an inosine residue

<220>
<221> misc_feature
<222> (1)..(25)
<223> Degenerate oligonucleotide Primer A wherein n represents inosine

<400> 7
arraygarra nggnrartay aarga 25

<210> 8
<211> 20

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: degenerate oligonucleotide PCR primer B wherein the letter "n" represents an inosine residue

<220>
<221> misc_feature
<222> (1)..(20)
<223> oligonucleotide PCR primer B wherein the letter n represents an inosine residue

<400> 8
atgytncary tntaygargc 20

<210> 9
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: degenerate oligonucleotide PCR primer C wherein the letter "n" represents an inosine residue

<220>
<221> misc_feature
<222> (1)..(24)
<223> PCR primer C wherein the letter n represents inosine

<400> 9
ctnkynrang gnctratrta ckty 24

<210> 10
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: degenerate oligonucleotide PCR primer D wherein the letter "n" represents an inosine residue

<220>
<221> misc_feature
<222> (1)..(23)
<223> PCR primer D wherein the letter n represents inosine

<400> 10
gaygaynnnt wygaygcnya ygg 23

<210> 11
<211> 108

<212> DNA
<213> Artificial Sequence

<400> 11
gatgatgggt ttgatgcgca cggaacccta gatgaattga agctattcac tgaggctgtg 60

agaagatggg acctctcctt tacagacaac ttccccgatt acatgaaa 108

<210> 12
<211> 104
<212> DNA
<213> Abies grandis

<400> 12
gacgacgggt atgatgcgca tggaacgatt gacgagcttg aactcttcac atctgcaatt 60

aagagatgga attcatcaga gatagacagc ttccccgact atat 104

<210> 13
<211> 105
<212> DNA
<213> Abies grandis

<400> 13
gatgatgggt atgatgcgta cggaacgttg gaagaaatca aaatcatgac agagggagtg 60

agacgatggg atctttcgtt gaccgcttnc cccgactata tgaaa 105

<210> 14
<211> 117
<212> DNA
<213> Abies grandis

<400> 14
gacgatgggt atgatgcgca tggaaccttg gaccaactca aaatctttac agagggagtg 60

agacgatggg atgtttcgtt ggtagaccac ttnccccgac tacatgcaat ctagacc 117

<210> 15
<211> 2424
<212> DNA
<213> Abies grandis

<220>
<221> CDS
<222> (2)..(2350)
<223> Clone AG1.28

<400> 15
g ggt tat gat ctt gtg cat tct ctt aaa tca cct tat att gat tct agt 49
Gly Tyr Asp Leu Val His Ser Leu Lys Ser Pro Tyr Ile Asp Ser Ser
1 5 10 15

tac aga gaa cgc gcg gag gtc ctt gtt agc gag att aaa gtg atg ctt 97
Tyr Arg Glu Arg Ala Glu Val Leu Val Ser Glu Ile Lys Val Met Leu
20 25 30

```
aat cca gct att aca gga gat gga gaa tca atg att act cca tct gct   145
Asn Pro Ala Ile Thr Gly Asp Gly Glu Ser Met Ile Thr Pro Ser Ala
         35                  40                  45

tat gac aca gca tgg gta gcg agg gtg ccc gcc att gat ggc tct gct   193
Tyr Asp Thr Ala Trp Val Ala Arg Val Pro Ala Ile Asp Gly Ser Ala
     50                  55                  60

cgc ccg caa ttt ccc caa aca gtt gac tgg att ttg aaa aac cag tta   241
Arg Pro Gln Phe Pro Gln Thr Val Asp Trp Ile Leu Lys Asn Gln Leu
 65                  70                  75                  80

aaa gat ggt tca tgg gga att cag tcc cac ttt ctg ctg tcc gac cgt   289
Lys Asp Gly Ser Trp Gly Ile Gln Ser His Phe Leu Leu Ser Asp Arg
                 85                  90                  95

ctt ctt gcc act ctt tct tgt gtt ctt gtg ctc ctt aaa tgg aac gtt   337
Leu Leu Ala Thr Leu Ser Cys Val Leu Val Leu Leu Lys Trp Asn Val
            100                 105                 110

ggg gat ctg caa gta gag cag gga att gaa ttc ata aag agc aat ctg   385
Gly Asp Leu Gln Val Glu Gln Gly Ile Glu Phe Ile Lys Ser Asn Leu
        115                 120                 125

gaa cta gta aag gat gaa acc gat caa gat agc ttg gta aca gac ttt   433
Glu Leu Val Lys Asp Glu Thr Asp Gln Asp Ser Leu Val Thr Asp Phe
    130                 135                 140

gag atc ata ttt cct tct ctg tta aga gaa gct caa tct ctg cgc ctc   481
Glu Ile Ile Phe Pro Ser Leu Leu Arg Glu Ala Gln Ser Leu Arg Leu
145                 150                 155                 160

gga ctt ccc tac gac ctg cct tat ata cat ctg ttg cag act aaa cgg   529
Gly Leu Pro Tyr Asp Leu Pro Tyr Ile His Leu Leu Gln Thr Lys Arg
                165                 170                 175

cag gaa aga tta gca aaa ctt tca agg gag gaa att tat gcg gtt ccg   577
Gln Glu Arg Leu Ala Lys Leu Ser Arg Glu Glu Ile Tyr Ala Val Pro
            180                 185                 190

tcg cca ttg ttg tat tct tta gag gga ata caa gat ata gtt gaa tgg   625
Ser Pro Leu Leu Tyr Ser Leu Glu Gly Ile Gln Asp Ile Val Glu Trp
        195                 200                 205

gaa cga ata atg gaa gtt caa agt cag gat ggg tct ttc tta agc tca   673
Glu Arg Ile Met Glu Val Gln Ser Gln Asp Gly Ser Phe Leu Ser Ser
    210                 215                 220

cct gct tct act gcc tgc gtt ttc atg cac aca gga gac gcg aaa tgc   721
Pro Ala Ser Thr Ala Cys Val Phe Met His Thr Gly Asp Ala Lys Cys
225                 230                 235                 240

ctt gaa ttc ttg aac agt gtg atg atc aag ttt gga aat ttt gtt ccc   769
Leu Glu Phe Leu Asn Ser Val Met Ile Lys Phe Gly Asn Phe Val Pro
                245                 250                 255

tgc ctg tat cct gtg gat ctg ctg gaa cgc ctg ttg atc gta gat aat   817
Cys Leu Tyr Pro Val Asp Leu Leu Glu Arg Leu Leu Ile Val Asp Asn
            260                 265                 270
```

```
att gta cgc ctt gga atc tat aga cac ttt gaa aag gaa atc aag gaa   865
Ile Val Arg Leu Gly Ile Tyr Arg His Phe Glu Lys Glu Ile Lys Glu
        275                 280                 285

gct ctt gat tat gtt tac agg cat tgg aac gaa aga gga att ggg tgg   913
Ala Leu Asp Tyr Val Tyr Arg His Trp Asn Glu Arg Gly Ile Gly Trp
    290                 295                 300

ggc aga cta aat ccc ata gca gat ctt gag acc act gct ttg gga ttt   961
Gly Arg Leu Asn Pro Ile Ala Asp Leu Glu Thr Thr Ala Leu Gly Phe
305                 310                 315                 320

cga ttg ctt cgg ctg cat agg tac aat gta tct cca gcc att ttt gac   1009
Arg Leu Leu Arg Leu His Arg Tyr Asn Val Ser Pro Ala Ile Phe Asp
                325                 330                 335

aac ttc aaa gat gcc aat ggg aaa ttc att tgc tcg acc ggt caa ttc   1057
Asn Phe Lys Asp Ala Asn Gly Lys Phe Ile Cys Ser Thr Gly Gln Phe
            340                 345                 350

aac aaa gat gta gca agc atg ctg aat ctt tat aga gct tcc cag ctc   1105
Asn Lys Asp Val Ala Ser Met Leu Asn Leu Tyr Arg Ala Ser Gln Leu
        355                 360                 365

gca ttt ccc gga gaa aac att ctt gat gaa gct aaa agc ttc gct act   1153
Ala Phe Pro Gly Glu Asn Ile Leu Asp Glu Ala Lys Ser Phe Ala Thr
    370                 375                 380

aaa tat ttg aga gaa gct ctt gag aaa agt gag act tcc agt gca tgg   1201
Lys Tyr Leu Arg Glu Ala Leu Glu Lys Ser Glu Thr Ser Ser Ala Trp
385                 390                 395                 400

aac aac aaa caa aac ctg agc caa gag atc aaa tac gcg ctg aag act   1249
Asn Asn Lys Gln Asn Leu Ser Gln Glu Ile Lys Tyr Ala Leu Lys Thr
                405                 410                 415

tct tgg cat gcc agt gtt ccg aga gtg gaa gca aag aga tac tgt caa   1297
Ser Trp His Ala Ser Val Pro Arg Val Glu Ala Lys Arg Tyr Cys Gln
            420                 425                 430

gtg tat cgc cca gat tat gca cgc ata gca aaa tgc gtt tac aag cta   1345
Val Tyr Arg Pro Asp Tyr Ala Arg Ile Ala Lys Cys Val Tyr Lys Leu
        435                 440                 445

ccc tac gtg aac aat gaa aag ttt tta gag ctg gga aaa tta gat ttc   1393
Pro Tyr Val Asn Asn Glu Lys Phe Leu Glu Leu Gly Lys Leu Asp Phe
    450                 455                 460

aac att atc cag tcc atc cac caa gaa gaa atg aag aat gtt acc agc   1441
Asn Ile Ile Gln Ser Ile His Gln Glu Glu Met Lys Asn Val Thr Ser
465                 470                 475                 480

tgg ttt aga gat tcg ggg ttg cca cta ttc acc ttc gct cgg gag agg   1489
Trp Phe Arg Asp Ser Gly Leu Pro Leu Phe Thr Phe Ala Arg Glu Arg
                485                 490                 495

ccg ctg gaa ttc tac ttc tta gta gcg gcg ggg acc tat gaa ccc cag   1537
Pro Leu Glu Phe Tyr Phe Leu Val Ala Ala Gly Thr Tyr Glu Pro Gln
            500                 505                 510
```

```
tat gcc aaa tgc agg ttc ctc ttt aca aaa gtg gca tgc ttg cag act   1585
Tyr Ala Lys Cys Arg Phe Leu Phe Thr Lys Val Ala Cys Leu Gln Thr
        515                 520                 525

gtt ctg gac gat atg tat gac act tat gga acc cta gat gaa ttg aag   1633
Val Leu Asp Asp Met Tyr Asp Thr Tyr Gly Thr Leu Asp Glu Leu Lys
    530                 535                 540

cta ttc act gag gct gtg aga aga tgg gac ctc tcc ttt aca gaa aac   1681
Leu Phe Thr Glu Ala Val Arg Arg Trp Asp Leu Ser Phe Thr Glu Asn
545                 550                 555                 560

ctt cca gac tat atg aaa cta tgt tac caa atc tat tat gac ata gtt   1729
Leu Pro Asp Tyr Met Lys Leu Cys Tyr Gln Ile Tyr Tyr Asp Ile Val
                565                 570                 575

cac gag gtg gct tgg gag gca gag aag gaa cag ggg cgt gaa ttg gtc   1777
His Glu Val Ala Trp Glu Ala Glu Lys Glu Gln Gly Arg Glu Leu Val
            580                 585                 590

agc ttt ttc aga aag gga tgg gag gat tat ctt ctg ggt tat tat gaa   1825
Ser Phe Phe Arg Lys Gly Trp Glu Asp Tyr Leu Leu Gly Tyr Tyr Glu
        595                 600                 605

gaa gct gaa tgg tta gct gct gag tat gtg cct acc ttg gac gag tac   1873
Glu Ala Glu Trp Leu Ala Ala Glu Tyr Val Pro Thr Leu Asp Glu Tyr
    610                 615                 620

ata aag aat gga atc aca tct atc ggc caa cgt ata ctt ctg ttg agt   1921
Ile Lys Asn Gly Ile Thr Ser Ile Gly Gln Arg Ile Leu Leu Leu Ser
625                 630                 635                 640

gga gtg ttg ata atg gat ggg caa ctc ctt tcg caa gag gca tta gag   1969
Gly Val Leu Ile Met Asp Gly Gln Leu Leu Ser Gln Glu Ala Leu Glu
                645                 650                 655

aaa gta gat tat cca gga aga cgt gtt ctc aca gag ctg aat agc ctc   2017
Lys Val Asp Tyr Pro Gly Arg Arg Val Leu Thr Glu Leu Asn Ser Leu
            660                 665                 670

att tcc cgc ctg gcg gat gac acg aag aca tat aaa gct gag aag gct   2065
Ile Ser Arg Leu Ala Asp Asp Thr Lys Thr Tyr Lys Ala Glu Lys Ala
        675                 680                 685

cgt gga gaa ttg gcg tcc agc att gaa tgt tac atg aaa gac cat cct   2113
Arg Gly Glu Leu Ala Ser Ser Ile Glu Cys Tyr Met Lys Asp His Pro
    690                 695                 700

gaa tgt aca gag gaa gag gct ctc gat cac atc tat agc att ctg gag   2161
Glu Cys Thr Glu Glu Glu Ala Leu Asp His Ile Tyr Ser Ile Leu Glu
705                 710                 715                 720

ccg gcg gtg aag gaa ctg aca aga gag ttt ctg aag ccc gac gac gtc   2209
Pro Ala Val Lys Glu Leu Thr Arg Glu Phe Leu Lys Pro Asp Asp Val
                725                 730                 735

cca ttc gcc tgc aag aag atg ctt ttc gag gag aca aga gtg acg atg   2257
Pro Phe Ala Cys Lys Lys Met Leu Phe Glu Glu Thr Arg Val Thr Met
            740                 745                 750
```

```
gtg ata ttc aag gat gga gat gga ttc ggt gtt tcc aaa tta gaa gtc   2305
Val Ile Phe Lys Asp Gly Asp Gly Phe Gly Val Ser Lys Leu Glu Val
        755                 760                 765

aaa gat cat atc aaa gag tgt ctc att gaa ccg ctg cca ctg taa       2350
Lys Asp His Ile Lys Glu Cys Leu Ile Glu Pro Leu Pro Leu
    770                 775                 780

tcaaaatagt tgcaataata attgaaataa tgtcaactat gtttcacaaa aaaaaaaaaa 2410

aaaaaaaaaa aaaa                                                   2424
```

<210> 16
<211> 782
<212> PRT
<213> Abies grandis

<400> 16

```
Gly Tyr Asp Leu Val His Ser Leu Lys Ser Pro Tyr Ile Asp Ser Ser
  1               5                  10                  15

Tyr Arg Glu Arg Ala Glu Val Leu Val Ser Glu Ile Lys Val Met Leu
            20                  25                  30

Asn Pro Ala Ile Thr Gly Asp Gly Glu Ser Met Ile Thr Pro Ser Ala
        35                  40                  45

Tyr Asp Thr Ala Trp Val Ala Arg Val Pro Ala Ile Asp Gly Ser Ala
    50                  55                  60

Arg Pro Gln Phe Pro Gln Thr Val Asp Trp Ile Leu Lys Asn Gln Leu
65                  70                  75                  80

Lys Asp Gly Ser Trp Gly Ile Gln Ser His Phe Leu Leu Ser Asp Arg
                85                  90                  95

Leu Leu Ala Thr Leu Ser Cys Val Leu Val Leu Leu Lys Trp Asn Val
            100                 105                 110

Gly Asp Leu Gln Val Glu Gln Gly Ile Glu Phe Ile Lys Ser Asn Leu
        115                 120                 125

Glu Leu Val Lys Asp Glu Thr Asp Gln Asp Ser Leu Val Thr Asp Phe
    130                 135                 140

Glu Ile Ile Phe Pro Ser Leu Leu Arg Glu Ala Gln Ser Leu Arg Leu
145                 150                 155                 160

Gly Leu Pro Tyr Asp Leu Pro Tyr Ile His Leu Leu Gln Thr Lys Arg
                165                 170                 175

Gln Glu Arg Leu Ala Lys Leu Ser Arg Glu Glu Ile Tyr Ala Val Pro
            180                 185                 190

Ser Pro Leu Leu Tyr Ser Leu Glu Gly Ile Gln Asp Ile Val Glu Trp
        195                 200                 205

Glu Arg Ile Met Glu Val Gln Ser Gln Asp Gly Ser Phe Leu Ser Ser
```

210 215 220

Pro Ala Ser Thr Ala Cys Val Phe Met His Thr Gly Asp Ala Lys Cys
225 230 235 240

Leu Glu Phe Leu Asn Ser Val Met Ile Lys Phe Gly Asn Phe Val Pro
245 250 255

Cys Leu Tyr Pro Val Asp Leu Leu Glu Arg Leu Leu Ile Val Asp Asn
260 265 270

Ile Val Arg Leu Gly Ile Tyr Arg His Phe Glu Lys Glu Ile Lys Glu
275 280 285

Ala Leu Asp Tyr Val Tyr Arg His Trp Asn Glu Arg Gly Ile Gly Trp
290 295 300

Gly Arg Leu Asn Pro Ile Ala Asp Leu Glu Thr Thr Ala Leu Gly Phe
305 310 315 320

Arg Leu Leu Arg Leu His Arg Tyr Asn Val Ser Pro Ala Ile Phe Asp
325 330 335

Asn Phe Lys Asp Ala Asn Gly Lys Phe Ile Cys Ser Thr Gly Gln Phe
340 345 350

Asn Lys Asp Val Ala Ser Met Leu Asn Leu Tyr Arg Ala Ser Gln Leu
355 360 365

Ala Phe Pro Gly Glu Asn Ile Leu Asp Glu Ala Lys Ser Phe Ala Thr
370 375 380

Lys Tyr Leu Arg Glu Ala Leu Glu Lys Ser Glu Thr Ser Ser Ala Trp
385 390 395 400

Asn Asn Lys Gln Asn Leu Ser Gln Glu Ile Lys Tyr Ala Leu Lys Thr
405 410 415

Ser Trp His Ala Ser Val Pro Arg Val Glu Ala Lys Arg Tyr Cys Gln
420 425 430

Val Tyr Arg Pro Asp Tyr Ala Arg Ile Ala Lys Cys Val Tyr Lys Leu
435 440 445

Pro Tyr Val Asn Asn Glu Lys Phe Leu Glu Leu Gly Lys Leu Asp Phe
450 455 460

Asn Ile Ile Gln Ser Ile His Gln Glu Glu Met Lys Asn Val Thr Ser
465 470 475 480

Trp Phe Arg Asp Ser Gly Leu Pro Leu Phe Thr Phe Ala Arg Glu Arg
485 490 495

Pro Leu Glu Phe Tyr Phe Leu Val Ala Ala Gly Thr Tyr Glu Pro Gln
500 505 510

Tyr Ala Lys Cys Arg Phe Leu Phe Thr Lys Val Ala Cys Leu Gln Thr
515 520 525

Val Leu Asp Asp Met Tyr Asp Thr Tyr Gly Thr Leu Asp Glu Leu Lys

```
    530                 535                 540

Leu Phe Thr Glu Ala Val Arg Arg Trp Asp Leu Ser Phe Thr Glu Asn
545                 550                 555                 560

Leu Pro Asp Tyr Met Lys Leu Cys Tyr Gln Ile Tyr Tyr Asp Ile Val
                565                 570                 575

His Glu Val Ala Trp Glu Ala Glu Lys Glu Gln Gly Arg Glu Leu Val
            580                 585                 590

Ser Phe Phe Arg Lys Gly Trp Glu Asp Tyr Leu Leu Gly Tyr Tyr Glu
        595                 600                 605

Glu Ala Glu Trp Leu Ala Ala Glu Tyr Val Pro Thr Leu Asp Glu Tyr
    610                 615                 620

Ile Lys Asn Gly Ile Thr Ser Ile Gly Gln Arg Ile Leu Leu Leu Ser
625                 630                 635                 640

Gly Val Leu Ile Met Asp Gly Gln Leu Leu Ser Gln Glu Ala Leu Glu
                645                 650                 655

Lys Val Asp Tyr Pro Gly Arg Arg Val Leu Thr Glu Leu Asn Ser Leu
            660                 665                 670

Ile Ser Arg Leu Ala Asp Asp Thr Lys Thr Tyr Lys Ala Glu Lys Ala
        675                 680                 685

Arg Gly Glu Leu Ala Ser Ser Ile Glu Cys Tyr Met Lys Asp His Pro
    690                 695                 700

Glu Cys Thr Glu Glu Glu Ala Leu Asp His Ile Tyr Ser Ile Leu Glu
705                 710                 715                 720

Pro Ala Val Lys Glu Leu Thr Arg Glu Phe Leu Lys Pro Asp Asp Val
                725                 730                 735

Pro Phe Ala Cys Lys Lys Met Leu Phe Glu Glu Thr Arg Val Thr Met
            740                 745                 750

Val Ile Phe Lys Asp Gly Asp Gly Phe Gly Val Ser Lys Leu Glu Val
        755                 760                 765

Lys Asp His Ile Lys Glu Cys Leu Ile Glu Pro Leu Pro Leu
    770                 775                 780


<210> 17
<211> 1967
<212> DNA
<213> Abies grandis

<220>
<221> CDS
<222> (3)..(1736)
<223> Clone AG4.30

<400> 17
tt tct gaa tct tcc atc cct cga cgc aca ggg aat cat cac gga aat    47
```

```
   Ser Glu Ser Ser Ile Pro Arg Arg Thr Gly Asn His His Gly Asn
     1               5                  10                  15

gtg tgg gac gat gac ctc ata cac tct ctc aac tcg ccc tat ggg gca    95
Val Trp Asp Asp Asp Leu Ile His Ser Leu Asn Ser Pro Tyr Gly Ala
                20                  25                  30

cct gca tat tat gag ctc ctt caa aag ctt att gag gag atc aag cat   143
Pro Ala Tyr Tyr Glu Leu Leu Gln Lys Leu Ile Glu Glu Ile Lys His
            35                  40                  45

tta ctt ttg act gaa atg gaa atg gat gat ggc gat cat gat tta atc   191
Leu Leu Leu Thr Glu Met Glu Met Asp Asp Gly Asp His Asp Leu Ile
        50                  55                  60

aaa cgt ctt cag atc gtt gac act ttg gaa tgc ctg gga atc gat aga   239
Lys Arg Leu Gln Ile Val Asp Thr Leu Glu Cys Leu Gly Ile Asp Arg
    65                  70                  75

cat ttt gaa cac gaa ata caa aca gct gct tta gat tac gtt tac aga   287
His Phe Glu His Glu Ile Gln Thr Ala Ala Leu Asp Tyr Val Tyr Arg
80                  85                  90                  95

tgg tgg aac gaa aaa ggt atc ggg gag gga tca aga gat tcc ttc agc   335
Trp Trp Asn Glu Lys Gly Ile Gly Glu Gly Ser Arg Asp Ser Phe Ser
                100                 105                 110

aaa gat ctc aac gct aca gct tta gga ttt cgc gct ctc cga ctg cat   383
Lys Asp Leu Asn Ala Thr Ala Leu Gly Phe Arg Ala Leu Arg Leu His
            115                 120                 125

cga tat aac gta tcg tca ggt gtg ttg aag aat ttc aag gat gaa aac   431
Arg Tyr Asn Val Ser Ser Gly Val Leu Lys Asn Phe Lys Asp Glu Asn
        130                 135                 140

ggg aag ttc ttc tgc aac ttt act ggt gaa gaa gga aga gga gat aaa   479
Gly Lys Phe Phe Cys Asn Phe Thr Gly Glu Glu Gly Arg Gly Asp Lys
    145                 150                 155

caa gtg aga agc atg ttg tcg tta ctt cga gct tca gag att tcg ttt   527
Gln Val Arg Ser Met Leu Ser Leu Leu Arg Ala Ser Glu Ile Ser Phe
160                 165                 170                 175

ccc gga gaa aaa gtg atg gaa gag gcc aag gca ttc aca aga gaa tat   575
Pro Gly Glu Lys Val Met Glu Glu Ala Lys Ala Phe Thr Arg Glu Tyr
                180                 185                 190

cta aac caa gtt tta gct gga cac ggg gat gtg act gac gtg gat caa   623
Leu Asn Gln Val Leu Ala Gly His Gly Asp Val Thr Asp Val Asp Gln
            195                 200                 205

agc ctt ttg gag aga ggt gaa gta cgc att gga gtt tcc atg gct tgc   671
Ser Leu Leu Glu Arg Gly Glu Val Arg Ile Gly Val Ser Met Ala Cys
        210                 215                 220

agt gtg ccg aga tgg gag gca agg agc ttt ctc gaa ata tat gga cac   719
Ser Val Pro Arg Trp Glu Ala Arg Ser Phe Leu Glu Ile Tyr Gly His
    225                 230                 235

aac cat tcg tgg ctc aag tcg aat atc aac caa aaa atg ttg aag tta   767
```

```
Asn His Ser Trp Leu Lys Ser Asn Ile Asn Gln Lys Met Leu Lys Leu
240                 245                 250                 255

gcc aaa ttg gac ttc aat att ctg caa tgc aaa cat cac aag gag ata   815
Ala Lys Leu Asp Phe Asn Ile Leu Gln Cys Lys His His Lys Glu Ile
                260                 265                 270

cag ttt att aca agg tgg tgg aga gac tcg ggt ata tcg cag ctg aat   863
Gln Phe Ile Thr Arg Trp Trp Arg Asp Ser Gly Ile Ser Gln Leu Asn
            275                 280                 285

ttc tat cga aag cga cac gtg gaa tat tat tct tgg gtt gtt atg tgc   911
Phe Tyr Arg Lys Arg His Val Glu Tyr Tyr Ser Trp Val Val Met Cys
        290                 295                 300

att ttt gag cca gag ttc tct gaa agt aga att gcc ttc gcc aaa act   959
Ile Phe Glu Pro Glu Phe Ser Glu Ser Arg Ile Ala Phe Ala Lys Thr
    305                 310                 315

gct atc cta tgt act gtt cta gat gac ctc tat gat acg cac gca acg  1007
Ala Ile Leu Cys Thr Val Leu Asp Asp Leu Tyr Asp Thr His Ala Thr
320                 325                 330                 335

ttg cat gaa atc aaa atc atg aca gag gga gtg aga cga tgg gat ctt  1055
Leu His Glu Ile Lys Ile Met Thr Glu Gly Val Arg Arg Trp Asp Leu
                340                 345                 350

tcg ttg aca gat gac ctc cca gac tac att aaa att gca ttc cag ttc  1103
Ser Leu Thr Asp Asp Leu Pro Asp Tyr Ile Lys Ile Ala Phe Gln Phe
            355                 360                 365

ttc ttc aat aca gtg aat gaa ttg ata gtt gaa atc gtg aaa cgg caa  1151
Phe Phe Asn Thr Val Asn Glu Leu Ile Val Glu Ile Val Lys Arg Gln
        370                 375                 380

ggg cgg gat atg aca acc ata gtt aaa gat tgc tgg aag cga tac att  1199
Gly Arg Asp Met Thr Thr Ile Val Lys Asp Cys Trp Lys Arg Tyr Ile
    385                 390                 395

gag tct tat ctg caa gaa gcg gaa tgg ata gca act gga cat att ccc  1247
Glu Ser Tyr Leu Gln Glu Ala Glu Trp Ile Ala Thr Gly His Ile Pro
400                 405                 410                 415

act ttt aac gaa tac ata aag aac ggc atg gct agc tca ggg atg tgt  1295
Thr Phe Asn Glu Tyr Ile Lys Asn Gly Met Ala Ser Ser Gly Met Cys
                420                 425                 430

att gta aat ttg aat cca ctt ctc ttg ttg ggt aaa ctt ctc ccc gac  1343
Ile Val Asn Leu Asn Pro Leu Leu Leu Leu Gly Lys Leu Leu Pro Asp
            435                 440                 445

aac att ctg gag caa ata cat tct cca tcc aag atc ctg gac ctc tta  1391
Asn Ile Leu Glu Gln Ile His Ser Pro Ser Lys Ile Leu Asp Leu Leu
        450                 455                 460

gaa ttg acg ggc aga atc gcc gat gac tta aaa gat ttc gag gac gag  1439
Glu Leu Thr Gly Arg Ile Ala Asp Asp Leu Lys Asp Phe Glu Asp Glu
    465                 470                 475

aag gaa cgc ggg gag atg gct tca tct tta cag tgt tat atg aaa gaa  1487
```

```
Lys Glu Arg Gly Glu Met Ala Ser Ser Leu Gln Cys Tyr Met Lys Glu
480                 485                 490                 495

aat cct gaa tct aca gtg gaa aat gct tta aat cac ata aaa ggc atc   1535
Asn Pro Glu Ser Thr Val Glu Asn Ala Leu Asn His Ile Lys Gly Ile
                500                 505                 510

ctt aat cgt tcc ctt gag gaa ttt aat tgg gag ttt atg aag cag gat   1583
Leu Asn Arg Ser Leu Glu Glu Phe Asn Trp Glu Phe Met Lys Gln Asp
            515                 520                 525

agt gtc cca atg tgt tgc aag aaa ttc act ttc aat ata ggt cga gga   1631
Ser Val Pro Met Cys Cys Lys Lys Phe Thr Phe Asn Ile Gly Arg Gly
        530                 535                 540

ctt caa ttc atc tac aaa tac aga gac ggc tta tac att tct gac aag   1679
Leu Gln Phe Ile Tyr Lys Tyr Arg Asp Gly Leu Tyr Ile Ser Asp Lys
    545                 550                 555

gaa gta aag gac cag ata ttc aaa att cta gtc cac caa gtt cca atg   1727
Glu Val Lys Asp Gln Ile Phe Lys Ile Leu Val His Gln Val Pro Met
560                 565                 570                 575

gag gaa tag tgatggtctt ggttgtagtt gtctattatg gtatattgca           1776
Glu Glu

ttgacattta tgcttaaagg tgtttcttaa acgtttaggg cggaccgtta aataagttgg 1836

caataattaa tatttagaga ctttgtggaa gtgtttaggg cataaaattg cctatggcct 1896

atggcaagct acaaattgaa attgttgtgt ttataatatt tttattttat ttaaaaaaaa 1956

aaaaaaaaaa a                                                      1967


<210> 18
<211> 577
<212> PRT
<213> Abies grandis

<400> 18
Ser Glu Ser Ser Ile Pro Arg Arg Thr Gly Asn His His Gly Asn Val
  1               5                  10                  15

Trp Asp Asp Asp Leu Ile His Ser Leu Asn Ser Pro Tyr Gly Ala Pro
             20                  25                  30

Ala Tyr Tyr Glu Leu Leu Gln Lys Leu Ile Glu Glu Ile Lys His Leu
         35                  40                  45

Leu Leu Thr Glu Met Glu Met Asp Asp Gly Asp His Asp Leu Ile Lys
     50                  55                  60

Arg Leu Gln Ile Val Asp Thr Leu Glu Cys Leu Gly Ile Asp Arg His
 65                  70                  75                  80

Phe Glu His Glu Ile Gln Thr Ala Ala Leu Asp Tyr Val Tyr Arg Trp
                 85                  90                  95

Trp Asn Glu Lys Gly Ile Gly Glu Gly Ser Arg Asp Ser Phe Ser Lys
```

100 105 110

Asp Leu Asn Ala Thr Ala Leu Gly Phe Arg Ala Leu Arg Leu His Arg
115 120 125

Tyr Asn Val Ser Ser Gly Val Leu Lys Asn Phe Lys Asp Glu Asn Gly
130 135 140

Lys Phe Phe Cys Asn Phe Thr Gly Glu Glu Gly Arg Gly Asp Lys Gln
145 150 155 160

Val Arg Ser Met Leu Ser Leu Leu Arg Ala Ser Glu Ile Ser Phe Pro
165 170 175

Gly Glu Lys Val Met Glu Glu Ala Lys Ala Phe Thr Arg Glu Tyr Leu
180 185 190

Asn Gln Val Leu Ala Gly His Gly Asp Val Thr Asp Val Asp Gln Ser
195 200 205

Leu Leu Glu Arg Gly Glu Val Arg Ile Gly Val Ser Met Ala Cys Ser
210 215 220

Val Pro Arg Trp Glu Ala Arg Ser Phe Leu Glu Ile Tyr Gly His Asn
225 230 235 240

His Ser Trp Leu Lys Ser Asn Ile Asn Gln Lys Met Leu Lys Leu Ala
245 250 255

Lys Leu Asp Phe Asn Ile Leu Gln Cys Lys His His Lys Glu Ile Gln
260 265 270

Phe Ile Thr Arg Trp Trp Arg Asp Ser Gly Ile Ser Gln Leu Asn Phe
275 280 285

Tyr Arg Lys Arg His Val Glu Tyr Tyr Ser Trp Val Val Met Cys Ile
290 295 300

Phe Glu Pro Glu Phe Ser Glu Ser Arg Ile Ala Phe Ala Lys Thr Ala
305 310 315 320

Ile Leu Cys Thr Val Leu Asp Asp Leu Tyr Asp Thr His Ala Thr Leu
325 330 335

His Glu Ile Lys Ile Met Thr Glu Gly Val Arg Arg Trp Asp Leu Ser
340 345 350

Leu Thr Asp Asp Leu Pro Asp Tyr Ile Lys Ile Ala Phe Gln Phe Phe
355 360 365

Phe Asn Thr Val Asn Glu Leu Ile Val Glu Ile Val Lys Arg Gln Gly
370 375 380

Arg Asp Met Thr Thr Ile Val Lys Asp Cys Trp Lys Arg Tyr Ile Glu
385 390 395 400

Ser Tyr Leu Gln Glu Ala Glu Trp Ile Ala Thr Gly His Ile Pro Thr
405 410 415

Phe Asn Glu Tyr Ile Lys Asn Gly Met Ala Ser Ser Gly Met Cys Ile

```
            420                 425                 430

Val Asn Leu Asn Pro Leu Leu Leu Leu Gly Lys Leu Leu Pro Asp Asn
        435                 440                 445

Ile Leu Glu Gln Ile His Ser Pro Ser Lys Ile Leu Asp Leu Leu Glu
    450                 455                 460

Leu Thr Gly Arg Ile Ala Asp Asp Leu Lys Asp Phe Glu Asp Glu Lys
465                 470                 475                 480

Glu Arg Gly Glu Met Ala Ser Ser Leu Gln Cys Tyr Met Lys Glu Asn
                485                 490                 495

Pro Glu Ser Thr Val Glu Asn Ala Leu Asn His Ile Lys Gly Ile Leu
            500                 505                 510

Asn Arg Ser Leu Glu Glu Phe Asn Trp Glu Phe Met Lys Gln Asp Ser
        515                 520                 525

Val Pro Met Cys Cys Lys Lys Phe Thr Phe Asn Ile Gly Arg Gly Leu
    530                 535                 540

Gln Phe Ile Tyr Lys Tyr Arg Asp Gly Leu Tyr Ile Ser Asp Lys Glu
545                 550                 555                 560

Val Lys Asp Gln Ile Phe Lys Ile Leu Val His Gln Val Pro Met Glu
                565                 570                 575

Glu
```

<210> 19
<211> 1416
<212> DNA
<213> Abies grandis

<220>
<221> CDS
<222> (3)..(1199)
<223> Clone AG5.9

<400> 19

```
aa aaa gtg atg gaa gag gcg aag gca ttc aca aca aat tat cta aag    47
   Lys Val Met Glu Glu Ala Lys Ala Phe Thr Thr Asn Tyr Leu Lys
     1               5                  10                  15

aaa gtt tta gca gga cgg gag gct acc cac gtc gat gaa agc ctt ttg    95
Lys Val Leu Ala Gly Arg Glu Ala Thr His Val Asp Glu Ser Leu Leu
                20                  25                  30

gga gag gtg aag tac gca ttg gag ttt cca tgg cat tgc agt gtg cag   143
Gly Glu Val Lys Tyr Ala Leu Glu Phe Pro Trp His Cys Ser Val Gln
            35                  40                  45

aga tgg gag gca agg agc ttt atc gaa ata ttt gga caa att gat tca   191
Arg Trp Glu Ala Arg Ser Phe Ile Glu Ile Phe Gly Gln Ile Asp Ser
        50                  55                  60

gag ctt aag tcg aat ttg agc aaa aaa atg tta gag ttg gcg aaa ttg   239
```

```
Glu Leu Lys Ser Asn Leu Ser Lys Lys Met Leu Glu Leu Ala Lys Leu
 65                  70                  75

gac ttc aat att ctg caa tgc aca cat cag aaa gaa ctg cag att atc   287
Asp Phe Asn Ile Leu Gln Cys Thr His Gln Lys Glu Leu Gln Ile Ile
 80                  85                  90                  95

tca agg tgg ttc gca gac tca agt ata gca tcc ctg aat ttc tat cgg   335
Ser Arg Trp Phe Ala Asp Ser Ser Ile Ala Ser Leu Asn Phe Tyr Arg
                100                 105                 110

aaa tgt tac gtc gaa ttt tac ttt tgg atg gct gca gcc atc tcc gag   383
Lys Cys Tyr Val Glu Phe Tyr Phe Trp Met Ala Ala Ala Ile Ser Glu
            115                 120                 125

ccg gag ttt tct gga agc aga gtt gcc ttc aca aaa att gct ata ctg   431
Pro Glu Phe Ser Gly Ser Arg Val Ala Phe Thr Lys Ile Ala Ile Leu
        130                 135                 140

atg aca atg cta gat gac ctg tac gat act cac gga acc ttg gac caa   479
Met Thr Met Leu Asp Asp Leu Tyr Asp Thr His Gly Thr Leu Asp Gln
    145                 150                 155

ctc aaa atc ttt aca gag gga gtg aga cga tgg gat gtt tcg ttg gta   527
Leu Lys Ile Phe Thr Glu Gly Val Arg Arg Trp Asp Val Ser Leu Val
160                 165                 170                 175

gag ggc ctc cca gac ttc atg aaa att gca ttc gag ttc tgg tta aag   575
Glu Gly Leu Pro Asp Phe Met Lys Ile Ala Phe Glu Phe Trp Leu Lys
                180                 185                 190

aca tct aat gaa ttg att gct gaa gct gtt aaa gcg caa ggg caa gat   623
Thr Ser Asn Glu Leu Ile Ala Glu Ala Val Lys Ala Gln Gly Gln Asp
            195                 200                 205

atg gcg gcc tac ata aga aaa aat gca tgg gag cga tac ctt gaa gct   671
Met Ala Ala Tyr Ile Arg Lys Asn Ala Trp Glu Arg Tyr Leu Glu Ala
        210                 215                 220

tat ctg caa gat gcg gaa tgg ata gcc act gga cat gtc ccc acc ttt   719
Tyr Leu Gln Asp Ala Glu Trp Ile Ala Thr Gly His Val Pro Thr Phe
    225                 230                 235

gat gag tac ttg aat aat ggc aca cca aac act ggg atg tgt gta ttg   767
Asp Glu Tyr Leu Asn Asn Gly Thr Pro Asn Thr Gly Met Cys Val Leu
240                 245                 250                 255

aat ttg att ccg ctt ctg tta atg ggt gaa cat tta cca atc gac att   815
Asn Leu Ile Pro Leu Leu Leu Met Gly Glu His Leu Pro Ile Asp Ile
                260                 265                 270

ctg gag caa ata ttc ttg ccc tcc agg ttc cac cat ctc att gaa ttg   863
Leu Glu Gln Ile Phe Leu Pro Ser Arg Phe His His Leu Ile Glu Leu
            275                 280                 285

gct tcc agg ctc gtc gat gac gcg aga gat ttc cag gcg gag aag gat   911
Ala Ser Arg Leu Val Asp Asp Ala Arg Asp Phe Gln Ala Glu Lys Asp
        290                 295                 300

cat ggg gat tta tcg tgt att gag tgt tat tta aaa gat cat cct gag   959
```

```
His Gly Asp Leu Ser Cys Ile Glu Cys Tyr Leu Lys Asp His Pro Glu
305                 310                 315

tct aca gta gaa gat gct tta aat cat gtt aat ggc ctc ctt ggc aat   1007
Ser Thr Val Glu Asp Ala Leu Asn His Val Asn Gly Leu Leu Gly Asn
320                 325                 330                 335

tgc ctt ctg gaa atg aat tgg aag ttc tta aag aag cag gac agt gtg   1055
Cys Leu Leu Glu Met Asn Trp Lys Phe Leu Lys Lys Gln Asp Ser Val
                340                 345                 350

cca ctc tcg tgt aag aag tac agc ttc cat gta ttg gca cga agc atc   1103
Pro Leu Ser Cys Lys Lys Tyr Ser Phe His Val Leu Ala Arg Ser Ile
            355                 360                 365

caa ttc atg tac aat caa ggc gat ggc ttc tcc att tcg aac aaa gtg   1151
Gln Phe Met Tyr Asn Gln Gly Asp Gly Phe Ser Ile Ser Asn Lys Val
        370                 375                 380

atc aag gat caa gtg cag aaa gtt ctt att gtc ccc gtg cct att tga   1199
Ile Lys Asp Gln Val Gln Lys Val Leu Ile Val Pro Val Pro Ile
385                 390                 395

tagtagatac tagatagtag attagtagct attagtattt atttcatatc aatatttact 1259

aatgctgatg atggttaaag tccattcaga ccaatctttg gtttattgga cttaaataaa 1319

tgaattaatt agtttgtttt aaaattgtac tatttactgt tggaaataat gttttcatta 1379

ttgaaataac tagcacaact attttagtgt ggttgat                          1416
```

<210> 20
<211> 398
<212> PRT
<213> Abies grandis

<400> 20

```
Lys Val Met Glu Glu Ala Lys Ala Phe Thr Thr Asn Tyr Leu Lys Lys
  1               5                  10                  15

Val Leu Ala Gly Arg Glu Ala Thr His Val Asp Glu Ser Leu Leu Gly
             20                  25                  30

Glu Val Lys Tyr Ala Leu Glu Phe Pro Trp His Cys Ser Val Gln Arg
         35                  40                  45

Trp Glu Ala Arg Ser Phe Ile Glu Ile Phe Gly Gln Ile Asp Ser Glu
     50                  55                  60

Leu Lys Ser Asn Leu Ser Lys Lys Met Leu Glu Leu Ala Lys Leu Asp
 65                  70                  75                  80

Phe Asn Ile Leu Gln Cys Thr His Gln Lys Glu Leu Gln Ile Ile Ser
                 85                  90                  95

Arg Trp Phe Ala Asp Ser Ser Ile Ala Ser Leu Asn Phe Tyr Arg Lys
            100                 105                 110

Cys Tyr Val Glu Phe Tyr Phe Trp Met Ala Ala Ala Ile Ser Glu Pro
```

```
        115                 120                 125

Glu Phe Ser Gly Ser Arg Val Ala Phe Thr Lys Ile Ala Ile Leu Met
    130                 135                 140

Thr Met Leu Asp Asp Leu Tyr Asp Thr His Gly Thr Leu Asp Gln Leu
145                 150                 155                 160

Lys Ile Phe Thr Glu Gly Val Arg Arg Trp Asp Val Ser Leu Val Glu
                165                 170                 175

Gly Leu Pro Asp Phe Met Lys Ile Ala Phe Glu Phe Trp Leu Lys Thr
            180                 185                 190

Ser Asn Glu Leu Ile Ala Glu Ala Val Lys Ala Gln Gly Gln Asp Met
        195                 200                 205

Ala Ala Tyr Ile Arg Lys Asn Ala Trp Glu Arg Tyr Leu Glu Ala Tyr
    210                 215                 220

Leu Gln Asp Ala Glu Trp Ile Ala Thr Gly His Val Pro Thr Phe Asp
225                 230                 235                 240

Glu Tyr Leu Asn Asn Gly Thr Pro Asn Thr Gly Met Cys Val Leu Asn
                245                 250                 255

Leu Ile Pro Leu Leu Leu Met Gly Glu His Leu Pro Ile Asp Ile Leu
            260                 265                 270

Glu Gln Ile Phe Leu Pro Ser Arg Phe His His Leu Ile Glu Leu Ala
        275                 280                 285

Ser Arg Leu Val Asp Asp Ala Arg Asp Phe Gln Ala Glu Lys Asp His
    290                 295                 300

Gly Asp Leu Ser Cys Ile Glu Cys Tyr Leu Lys Asp His Pro Glu Ser
305                 310                 315                 320

Thr Val Glu Asp Ala Leu Asn His Val Asn Gly Leu Leu Gly Asn Cys
                325                 330                 335

Leu Leu Glu Met Asn Trp Lys Phe Leu Lys Lys Gln Asp Ser Val Pro
            340                 345                 350

Leu Ser Cys Lys Lys Tyr Ser Phe His Val Leu Ala Arg Ser Ile Gln
        355                 360                 365

Phe Met Tyr Asn Gln Gly Asp Gly Phe Ser Ile Ser Asn Lys Val Ile
    370                 375                 380

Lys Asp Gln Val Gln Lys Val Leu Ile Val Pro Val Pro Ile
385                 390                 395
```

<210> 21
<211> 23
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:
oligonucleotide PCR primer E wherein the letter
"n" represents an inosine residue

<220>
<221> misc_feature
<222> (1)..(23)
<223> PCR primer E wherein the letter n represents
inosine

<400> 21
ggngaramrr tnatggarga rgc 23

<210> 22
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: degenerate
oligonucleotide primer F wherein the letter "n"
represents an inosine residue

<220>
<221> misc_feature
<222> (1)..(24)
<223> PCR primer F wherein the letter n represents
inosine

<400> 22
garytncary tnhbnmgntg gtgg 24

<210> 23
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: degenerate
oligonucleotide PCR primer G wherein the letter
"n" represents an inosine residue

<220>
<221> misc_feature
<222> (1)..(21)
<223> PCR primer G wherein the letter n represents
inosine

<400> 23
ccarttnarn ccyttnacrt c 21

<210> 24
<211> 533
<212> DNA
<213> Abies grandis

<400> 24

ggggaaaaaa tgatggagga agctgaaatc ttctctacca aatatttaaa agaagccctg 60

caaaagattc cggtctccag tctttcgcga gagatcgggg acgttttgga atatggttgg 120

cacacatatt tgccgcgatt ggaagcaagg aattacatcc aagtctttgg acaggacact 180

gagaacacga agtcatatgt gaagagcaaa aaacttttag aactcgcaaa attggagttc 240

aacatctttc aatccttact cgcatatccg cattgcaacc cattctgaca atggacatcc 300

cctttcctga tcatatcctc aaggaagttg acttcccatc aaagcttaac gacttggcat 360

gtgccatcct tcgattacga ggtgatacgc ggtgctacaa ggcggacagg gctcgtggag 420

aagaagcttc ctctatatca tgttatatga aagacaatcc tggagtatca gaggaagatg 480

ctctcgatca tatcaacgcc atgatcagtg acgaagtcaa aggcttcaat tgg 533

<210> 25
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: conserved amino acid motif on which the sequence of Primer D was based, wherein Xaa at position number 3 represents Thr or Ile, Xaa at position number 4 represents Ile or Tyr or Phe, Xaa at position number 6 represents Ala or Val and Xaa at position number 8 represents Ala or Gly

<220>
<221> SITE
<222> (1)..(8)
<223> conserved amino acid motif on which sequence of primer D was based

<400> 25
Asp Asp Xaa Xaa Asp Xaa Tyr Xaa
1 5

<210> 26
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: conserved amino acid motif on which the sequence of Primer E was based wherein Xaa at position 3 represents Lys or Thr, Xaa at position 4 represents Val or Ile, Xaa at position 6 represents Glu or Asp

<220>
<221> SITE
<222> (1)..(8)
<223> conserved amino acid sequence on which the sequence of primer E was based

<400> 26

Gly Glu Xaa Xaa Met Xaa Glu Ala
1 5

<210> 27
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: conserved amino acid sequence on which the sequence of primer F was based wherein Xaa at position 2 represents Phe or Tyr or Asp Xaa at position 3 represents Ile or Leu, Xaa at position 4 represents Thr or Leu or Arg

<220>
<221> SITE
<222> (1)..(7)
<223> conserved amino acid sequence on which the sequence of primer F was based

<400> 27
Gln Xaa Xaa Xaa Arg Trp Trp
1 5

<210> 28
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: conserved amino acid motif on which the sequence of primer G was based wherein Xaa at position 6 represents Phe or Leu

<220>
<221> SITE
<222> (1)..(8)
<223> conserved amino acid sequence on which the sequence of primer G was based

<400> 28
Asp Val Ile Lys Gly Xaa Asn Trp
1 5

<210> 29
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: T3 primer oligonucleotide sequence

<400> 29
aattaaccct cactaaaggg 20

<210> 30
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: T7
oligonucleotide primer sequence

<400> 30
gtaatacgac tcactatagg gc 22

<210> 31
<211> 2205
<212> DNA
<213> Abies grandis

<220>
<221> CDS
<222> (57)..(1943)
<223> Clone AG3.48

<400> 31
gttatcttga gcttcctcca tataggccaa cacatatcat atcaaaggga gcaaga atg 59
Met
1

gct ctg gtt tct atc tca ccg ttg gct tcg aaa tct tgc ctg cgc aag 107
Ala Leu Val Ser Ile Ser Pro Leu Ala Ser Lys Ser Cys Leu Arg Lys
5 10 15

tcg ttg atc agt tca att cat gaa cat aag cct ccc tat aga aca atc 155
Ser Leu Ile Ser Ser Ile His Glu His Lys Pro Pro Tyr Arg Thr Ile
20 25 30

cca aat ctt gga atg cgt agg cga ggg aaa tct gtc acg cct tcc atg 203
Pro Asn Leu Gly Met Arg Arg Arg Gly Lys Ser Val Thr Pro Ser Met
35 40 45

agc atc agt ttg gcc acc gct gca cct gat gat ggt gta caa aga cgc 251
Ser Ile Ser Leu Ala Thr Ala Ala Pro Asp Asp Gly Val Gln Arg Arg
50 55 60 65

ata ggt gac tac cat tcc aat atc tgg gac gat gat ttc ata cag tct 299
Ile Gly Asp Tyr His Ser Asn Ile Trp Asp Asp Asp Phe Ile Gln Ser
70 75 80

cta tca acg cat tat ggg gaa ccc tct tac cag gaa cgt gct gag aga 347
Leu Ser Thr His Tyr Gly Glu Pro Ser Tyr Gln Glu Arg Ala Glu Arg
85 90 95

tta att gtg gag gta aag aag ata ttc aat tca atg tac ctg gat gat 395
Leu Ile Val Glu Val Lys Lys Ile Phe Asn Ser Met Tyr Leu Asp Asp
100 105 110

gga aga tta atg agt tcc ttt aat gat ctc atg caa cgc ctt tgg ata 443
Gly Arg Leu Met Ser Ser Phe Asn Asp Leu Met Gln Arg Leu Trp Ile
115 120 125

```
gtc gat agc gtt gaa cgt ttg ggg ata gct aga cat ttc aag aac gag   491
Val Asp Ser Val Glu Arg Leu Gly Ile Ala Arg His Phe Lys Asn Glu
130                 135                 140                 145

ata aca tca gct ctg gat tat gtt ttc cgt tac tgg gag gaa aac ggc   539
Ile Thr Ser Ala Leu Asp Tyr Val Phe Arg Tyr Trp Glu Glu Asn Gly
                150                 155                 160

att gga tgt ggg aga gac agt att gtt act gat ctc aac tca act gcg   587
Ile Gly Cys Gly Arg Asp Ser Ile Val Thr Asp Leu Asn Ser Thr Ala
            165                 170                 175

ttg ggg ttt cga act ctt cga tta cac ggg tac act gta tct cca gag   635
Leu Gly Phe Arg Thr Leu Arg Leu His Gly Tyr Thr Val Ser Pro Glu
        180                 185                 190

gtt tta aaa gct ttt caa gat caa aat gga cag ttt gta tgc tcc ccc   683
Val Leu Lys Ala Phe Gln Asp Gln Asn Gly Gln Phe Val Cys Ser Pro
    195                 200                 205

ggt cag aca gag ggt gag atc aga agc gtt ctt aac tta tat cgg gct   731
Gly Gln Thr Glu Gly Glu Ile Arg Ser Val Leu Asn Leu Tyr Arg Ala
210                 215                 220                 225

tcc ctc att gcc ttc cct ggt gag aaa gtt atg gaa gaa gct gaa atc   779
Ser Leu Ile Ala Phe Pro Gly Glu Lys Val Met Glu Glu Ala Glu Ile
                230                 235                 240

ttc tcc aca aga tat ttg aaa gaa gct cta caa aag att cca gtc tcc   827
Phe Ser Thr Arg Tyr Leu Lys Glu Ala Leu Gln Lys Ile Pro Val Ser
            245                 250                 255

gct ctt tca caa gag ata aag ttt gtt atg gaa tat ggc tgg cac aca   875
Ala Leu Ser Gln Glu Ile Lys Phe Val Met Glu Tyr Gly Trp His Thr
        260                 265                 270

aat ttg cca aga ttg gaa gca aga aat tac ata gac aca ctt gag aaa   923
Asn Leu Pro Arg Leu Glu Ala Arg Asn Tyr Ile Asp Thr Leu Glu Lys
    275                 280                 285

gac acc agt gca tgg ctc aat aaa aat gct ggg aag aag ctt tta gaa   971
Asp Thr Ser Ala Trp Leu Asn Lys Asn Ala Gly Lys Lys Leu Leu Glu
290                 295                 300                 305

ctt gca aaa ttg gag ttc aat ata ttt aac tcc tta caa caa aag gaa   1019
Leu Ala Lys Leu Glu Phe Asn Ile Phe Asn Ser Leu Gln Gln Lys Glu
                310                 315                 320

tta caa tat ctt ttg aga tgg tgg aaa gag tcg gat ttg cct aaa ttg   1067
Leu Gln Tyr Leu Leu Arg Trp Trp Lys Glu Ser Asp Leu Pro Lys Leu
            325                 330                 335

aca ttt gct cgg cat cgt cat gtg gaa ttc tac act ttg gcc tct tgt   1115
Thr Phe Ala Arg His Arg His Val Glu Phe Tyr Thr Leu Ala Ser Cys
        340                 345                 350

att gcc att gac cca aaa cat tct gca ttc aga cta ggc ttc gcc aaa   1163
Ile Ala Ile Asp Pro Lys His Ser Ala Phe Arg Leu Gly Phe Ala Lys
    355                 360                 365
```

```
atg tgt cat ctt gtc aca gtt ttg gac gat att tac gac act ttt gga   1211
Met Cys His Leu Val Thr Val Leu Asp Asp Ile Tyr Asp Thr Phe Gly
370                 375                 380                 385

acg att gac gag ctt gaa ctc ttc aca tct gca att aag aga tgg aat   1259
Thr Ile Asp Glu Leu Glu Leu Phe Thr Ser Ala Ile Lys Arg Trp Asn
                390                 395                 400

tca tca gag ata gaa cac ctt cca gaa tat atg aaa tgt gtg tac atg   1307
Ser Ser Glu Ile Glu His Leu Pro Glu Tyr Met Lys Cys Val Tyr Met
            405                 410                 415

gtc gtg ttt gaa act gta aat gaa ctg aca cga gag gcg gag aag act   1355
Val Val Phe Glu Thr Val Asn Glu Leu Thr Arg Glu Ala Glu Lys Thr
        420                 425                 430

caa ggg aga aac act ctc aac tat gtt cga aag gct tgg gag gct tat   1403
Gln Gly Arg Asn Thr Leu Asn Tyr Val Arg Lys Ala Trp Glu Ala Tyr
    435                 440                 445

ttt gat tca tat atg gaa gaa gca aaa tgg atc tct aat ggt tat ctg   1451
Phe Asp Ser Tyr Met Glu Glu Ala Lys Trp Ile Ser Asn Gly Tyr Leu
450                 455                 460                 465

cca acg ttt gaa gag tac cat gag aat ggg aaa gtg agc tct gca tat   1499
Pro Thr Phe Glu Glu Tyr His Glu Asn Gly Lys Val Ser Ser Ala Tyr
                470                 475                 480

cgc gta gca aca ttg caa ccc atc ctc act ttg aat gca tgg ctt cct   1547
Arg Val Ala Thr Leu Gln Pro Ile Leu Thr Leu Asn Ala Trp Leu Pro
            485                 490                 495

gat tac atc ttg aag gga att gat ttt cca tcc agg ttc aat gat ttg   1595
Asp Tyr Ile Leu Lys Gly Ile Asp Phe Pro Ser Arg Phe Asn Asp Leu
        500                 505                 510

gca tcg tcc ttc ctt cgg cta cga ggt gac aca cgc tgc tac aag gcc   1643
Ala Ser Ser Phe Leu Arg Leu Arg Gly Asp Thr Arg Cys Tyr Lys Ala
    515                 520                 525

gat agg gat cgt ggt gaa gaa gct tcg tgt ata tca tgt tat atg aaa   1691
Asp Arg Asp Arg Gly Glu Glu Ala Ser Cys Ile Ser Cys Tyr Met Lys
530                 535                 540                 545

gac aat cct gga tca acc gaa gaa gat gcc ctc aat cat atc aat gcc   1739
Asp Asn Pro Gly Ser Thr Glu Glu Asp Ala Leu Asn His Ile Asn Ala
                550                 555                 560

atg gtc aat gac ata atc aaa gaa tta aat tgg gaa ctt cta aga tcc   1787
Met Val Asn Asp Ile Ile Lys Glu Leu Asn Trp Glu Leu Leu Arg Ser
            565                 570                 575

aac gac aat att cca atg ctg gcc aag aaa cat gct ttt gac ata aca   1835
Asn Asp Asn Ile Pro Met Leu Ala Lys Lys His Ala Phe Asp Ile Thr
        580                 585                 590

aga gct ctc cac cat ctc tac ata tat cga gat ggc ttt agt gtt gcc   1883
Arg Ala Leu His His Leu Tyr Ile Tyr Arg Asp Gly Phe Ser Val Ala
    595                 600                 605
```

```
aac aag gaa aca aaa aaa ttg gtt atg gaa aca ctc ctt gaa tct atg   1931
Asn Lys Glu Thr Lys Lys Leu Val Met Glu Thr Leu Leu Glu Ser Met
610                 615                 620                 625

ctt ttt taa cta taaccatatc cataataata agctcataat gctaaattat       1983
Leu Phe

tggccttatg acatagttta tgtatgtact tgtgtgaatt caatcatatc gtgtgggtat 2043

gattaaaaag ctagagctta ctaggttagt aacatggtga taaaagttat aaaatgtgag 2103

ttatagagat acccatgttg aataatgaat tacaaaaaga gaaatttatg tagaataaga 2163

ttggaagctt ttcaattgtt ttaaaaaaaa aaaaaaaaaa aa                    2205
```

<210> 32
<211> 627
<212> PRT
<213> Abies grandis

<400> 32

```
Met Ala Leu Val Ser Ile Ser Pro Leu Ala Ser Lys Ser Cys Leu Arg
  1               5                  10                  15

Lys Ser Leu Ile Ser Ser Ile His Glu His Lys Pro Pro Tyr Arg Thr
            20                  25                  30

Ile Pro Asn Leu Gly Met Arg Arg Arg Gly Lys Ser Val Thr Pro Ser
        35                  40                  45

Met Ser Ile Ser Leu Ala Thr Ala Ala Pro Asp Asp Gly Val Gln Arg
    50                  55                  60

Arg Ile Gly Asp Tyr His Ser Asn Ile Trp Asp Asp Asp Phe Ile Gln
65                  70                  75                  80

Ser Leu Ser Thr His Tyr Gly Glu Pro Ser Tyr Gln Glu Arg Ala Glu
                85                  90                  95

Arg Leu Ile Val Glu Val Lys Lys Ile Phe Asn Ser Met Tyr Leu Asp
            100                 105                 110

Asp Gly Arg Leu Met Ser Ser Phe Asn Asp Leu Met Gln Arg Leu Trp
        115                 120                 125

Ile Val Asp Ser Val Glu Arg Leu Gly Ile Ala Arg His Phe Lys Asn
    130                 135                 140

Glu Ile Thr Ser Ala Leu Asp Tyr Val Phe Arg Tyr Trp Glu Glu Asn
145                 150                 155                 160

Gly Ile Gly Cys Gly Arg Asp Ser Ile Val Thr Asp Leu Asn Ser Thr
                165                 170                 175

Ala Leu Gly Phe Arg Thr Leu Arg Leu His Gly Tyr Thr Val Ser Pro
            180                 185                 190

Glu Val Leu Lys Ala Phe Gln Asp Gln Asn Gly Gln Phe Val Cys Ser
```

195 200 205

Pro Gly Gln Thr Glu Gly Glu Ile Arg Ser Val Leu Asn Leu Tyr Arg
210 215 220

Ala Ser Leu Ile Ala Phe Pro Gly Glu Lys Val Met Glu Glu Ala Glu
225 230 235 240

Ile Phe Ser Thr Arg Tyr Leu Lys Glu Ala Leu Gln Lys Ile Pro Val
245 250 255

Ser Ala Leu Ser Gln Glu Ile Lys Phe Val Met Glu Tyr Gly Trp His
260 265 270

Thr Asn Leu Pro Arg Leu Glu Ala Arg Asn Tyr Ile Asp Thr Leu Glu
275 280 285

Lys Asp Thr Ser Ala Trp Leu Asn Lys Asn Ala Gly Lys Lys Leu Leu
290 295 300

Glu Leu Ala Lys Leu Glu Phe Asn Ile Phe Asn Ser Leu Gln Gln Lys
305 310 315 320

Glu Leu Gln Tyr Leu Leu Arg Trp Trp Lys Glu Ser Asp Leu Pro Lys
325 330 335

Leu Thr Phe Ala Arg His Arg His Val Glu Phe Tyr Thr Leu Ala Ser
340 345 350

Cys Ile Ala Ile Asp Pro Lys His Ser Ala Phe Arg Leu Gly Phe Ala
355 360 365

Lys Met Cys His Leu Val Thr Val Leu Asp Asp Ile Tyr Asp Thr Phe
370 375 380

Gly Thr Ile Asp Glu Leu Glu Leu Phe Thr Ser Ala Ile Lys Arg Trp
385 390 395 400

Asn Ser Ser Glu Ile Glu His Leu Pro Glu Tyr Met Lys Cys Val Tyr
405 410 415

Met Val Val Phe Glu Thr Val Asn Glu Leu Thr Arg Glu Ala Glu Lys
420 425 430

Thr Gln Gly Arg Asn Thr Leu Asn Tyr Val Arg Lys Ala Trp Glu Ala
435 440 445

Tyr Phe Asp Ser Tyr Met Glu Glu Ala Lys Trp Ile Ser Asn Gly Tyr
450 455 460

Leu Pro Thr Phe Glu Glu Tyr His Glu Asn Gly Lys Val Ser Ser Ala
465 470 475 480

Tyr Arg Val Ala Thr Leu Gln Pro Ile Leu Thr Leu Asn Ala Trp Leu
485 490 495

Pro Asp Tyr Ile Leu Lys Gly Ile Asp Phe Pro Ser Arg Phe Asn Asp
500 505 510

Leu Ala Ser Ser Phe Leu Arg Leu Arg Gly Asp Thr Arg Cys Tyr Lys

515 520 525

Ala Asp Arg Asp Arg Gly Glu Glu Ala Ser Cys Ile Ser Cys Tyr Met
530 535 540

Lys Asp Asn Pro Gly Ser Thr Glu Glu Asp Ala Leu Asn His Ile Asn
545 550 555 560

Ala Met Val Asn Asp Ile Ile Lys Glu Leu Asn Trp Glu Leu Leu Arg
565 570 575

Ser Asn Asp Asn Ile Pro Met Leu Ala Lys Lys His Ala Phe Asp Ile
580 585 590

Thr Arg Ala Leu His His Leu Tyr Ile Tyr Arg Asp Gly Phe Ser Val
595 600 605

Ala Asn Lys Glu Thr Lys Lys Leu Val Met Glu Thr Leu Leu Glu Ser
610 615 620

Met Leu Phe
625

<210> 33
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR
oligonucleotide primer 2.2 BamHI

<400> 33
caaagggatc cagaatggct ctgg 24

<210> 34
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR
oligonucleotide primer 2.2 Not I

<400> 34
agtaagcggc cgctttttaa tcatacccac 30

<210> 35
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR
oligonucleotide primer 3.18 EcoRI

<400> 35

ctgcaggaat tcggcacgag c 21

<210> 36
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR oligonucleotide primer 3.18 SmaI

<400> 36
catagccccg ggcatagatt tgagctg 27

<210> 37
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR oligonucleotide primer 10 NdeI

<400> 37
ggcaggaaca tatggctctc ctttctatcg 30

<210> 38
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR oligonucleotide primer 10 BamHI

<400> 38
tctagaacta gtggatcccc cgggctgcag 30

<210> 39
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR oligonucleotide primer JB29

<400> 39
ctaccattcc aatatctg 18

<210> 40
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR oligonucleotide primer 2-8

<400> 40
gttggatctt agaagttccc 20

<210> 41
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR oligonucleotide primer 3-9

<400> 41
tttccattcc aacctctggg 20

<210> 42
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR oligonucleotide primer 3-11

<400> 42
cgtaatggaa agctctggcg 20

<210> 43
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR oligonucleotide primer 7-1

<400> 43
ccttacacgc ctttggatgg 20

<210> 44
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR oligonucleotide sequence 7-3

<400> 44
tctgttgatc caggatggtc 20

<210> 45
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: conserved amino acid motif common to all prenyl transferases wherein Xaa at position 3 and 4 represents any amino acid

<400> 45
Asp Asp Xaa Xaa Asp
1 5

<210> 46
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: amino acid motif from which oligonucleotide primers can be synthesized that hybridize to the monoterpene synthases of the present invention, wherein Xaa at position 4 represents Leu or Ile or Val

<400> 46
His Ser Asn Xaa Trp Asp Asp Asp
1 5

<210> 47
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: amino acid motif from which degenerate oligonucleotides can be constructed that hybridize to the monoterpene synthases of the present invention

<400> 47
Ala Leu Asp Tyr Val Tyr
1 5

<210> 48
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: amino acid motif from which degenerate oligonucleotide sequences can be constructed that hybridize to the monoterpene synthases of the present invention

<400> 48

Glu Leu Ala Lys Leu Glu Phe
1 5

<210> 49
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: amino acid motif from which degenerate oligonucleotide sequences can be constructed that hybridize to monoterpene synthase clones of the present invention

<400> 49
Arg Trp Trp Lys Glu Ser
1 5

<210> 50
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: amino acid motif from which oligonucleotide sequeences can be constructed that hybridize to monoterpene synthase clones of the present invention, wherein Xaa at position 1 represents Val or Ile or Leu

<400> 50
Xaa Leu Asp Asp Met Tyr Asp
1 5

<210> 51
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: amino acid motif from which oligonucleotide sequences can be constructed that hybridize to monoterpene synthase clones of the present invention wherein Xaa at position 1 reperesents Val or Ile or Leu

<400> 51
Xaa Leu Asp Asp Leu Tyr Asp
1 5

<210> 52
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: amino acid motif from which oligonucleotide sequences can be constructed that hybridize to the monoterpene synthase clones of the present invention, wherein Xaa at position 1 represents Val or Ile or Leu

<400> 52
Xaa Leu Asp Asp Ile Tyr Asp
1 5

<210> 53
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: amino acid motif from which oligonucleotide sequences can be constructed that hybridize to the monoterpene synthase clones of the present invention, wherein Xaa at position 6 represents Asn or His

<400> 53
Cys Tyr Met Lys Asp Xaa Pro
1 5

<210> 54
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: exemplary oligonucleotide that corresponds to peptide sequence MetMetMet

<400> 54
atgatgatg 9

<210> 55
<211> 9
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: exemplary oligonucleotide sequence that corresponds to peptide sequence MetMetMet

<400> 55
tactactac 9

<210> 56
<211> 9
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: exemplary oligonucleotide that corresponds to peptide sequence MetMetMet, n is inosine

<400> 56
nacnacnac 9

<210> 57
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: oligonucleotide corresponding to amino acid sequence set forth in SEQ ID NO:46

<220>
<221> misc_feature
<222> (1)..(24)
<223> Oligonucleotide that corresponds to the conserved amino acid sequence set forth in SEQ ID NO:46

<400> 57
gtgtcgttgg agaccctgct gctg 24

<210> 58
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: oligonucleotide sequence corresponding to amino acid sequence set forth in SEQ ID NO:47

<220>
<221> misc_feature
<222> (1)..(18)
<223> Oligonucleotide corresponding to amino acid sequence set forth in SEQ ID NO:47

<400> 58
cgggagctga tgcagatg 18

<210> 59
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: oligonucleotide that corresponds to amino acid sequence set forth in SEQ ID NO:48

<220>
<221> misc_feature
<222> (1)..(21)
<223> Oligonucleotide that corresponds to conserved
amino acid sequence set forth in SEQ ID NO:48

<400> 59
ctcgagcggt tcgagctcaa g 21

<210> 60
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
oligonucleotide that corresponds to amino acid
sequence set forth in SEQ ID NO:49

<220>
<221> misc_feature
<222> (1)..(18)
<223> Oligonucleotide that corresponds to conserved
amino acid sequence set forth in SEQ ID NO:49

<400> 60
gccaccacct tcctctcg 18

<210> 61
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
oligonucleotide sequence corresponding to amino
acid sequence set forth in SEQ ID NO:50

<220>
<221> misc_feature
<222> (1)..(21)
<223> Oligonucleotide sequence corresponding to amino
acid sequence set forth in SEQ ID NO:50

<400> 61
gaggagctgc tgtacatgct g 21

<210> 62
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
oligonucleotide corresponding to amino acid
sequence set forth in SEQ ID NO:51

<220>
<221> misc_feature
<222> (1)..(21)
<223> Oligonucleotide corresponding to conserved amino
acid sequence set forth in SEQ ID NO:51

<400> 62
gaggagctgc tggagatgct g 21

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/14528

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :426/601, 622; 435/320.1, 419; 530/370; 536/23.6; 800/281, 298
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 426/601, 622; 435/320.1, 419; 530/370; 536/23.6; 800/281, 298

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 95/11913 A1 (WASHINGTON STATE UNIVERSITY RESEARCH FOUNDATION) 04 May 1995, see page 3, lines 13-35 and pages 32-41.	1,3,5,7,10,13, 15-22,24,26-46, 49,50
Y	WILDUNG et al. A cDNA Clone for Taxadiene Synthase, the Diterpene Cyclase That Catalyzes the Committed Step of Taxol Biosynthesis. The Journal of Biological Chemistry. 19 April 1996. Vol. 271, No. 16, pages 9201-9204, see especially pages 9201-9202.	1,3,5,7,10, 13, 15-22,24,26-46, 49,50

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 26 OCTOBER 1998	Date of mailing of the international search report 10 NOV 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer THANDA WAI Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1998)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/14528

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	COLBY et al. 4S-Limonene Synthase from the Oil Glands of Spearmint (Mentha spicata). The Journal of Biological Chemistry. 05 November 1993. Vol. 268, No. 31, pages 23016-23024, see especially pages 23017-23018.	1,3,5,7,10, 13, 15-22,24,26-46, 49,50
Y	YUBA et al. cDNA Cloning, Characterization, and Functional Expression of 4S(-)-Limonene Synthase from Perilla frutescens. Archives of Biochemistry and Biophysics. 15 August 1996. Vol. 332, No. 21, pages 280287, see especially pages 281-282.	1,3,5,7,10,13, 15-22,24,26-46, 49,50
A,P	BOHLMANN et al. Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (E)-alpha-bisabolene synthase from grand fir (Abies grandis). PNAS USA. June 1998. Vol 95, No. 12, pages 6756-6761, see entire document.	1-46,49,50
A, P	BOHLMANN et al. Monoterpene Synthases from Grand Fir (Abies grandis): cDNA Isolation, Characterization, and Functional Expression of Myrcene Synthase, (-)-(4S)-Limonene Synthase, and (-)-(1S,5S)-Pinene Synthase. The Journal of Biological Chemistry. 29 August 1997, Vol. 272, No. 35, pages 21784-21792, see entire document.	1-50
Y	US 4,948,811 A (SPINNER et al.) 14 August 1990) see column 2, lines 20-54 and column 3, lines 57-50.	47, 48

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/14528

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

A01H 5/00, 5/10; A23D 7/00; A23K 1/14; C07H 21/04; C07K 4/10; C12N 15/04, 15/63, 15/82

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, AGRICOLA, BIOSIS, EMBASE, WPIDS, MPSEARCH (for SEQ ID NO:1-6 and 46-53)

search terms: monoterpene synthase#, myrcene synthase#, pinene synthase#, gymnosperm?, grand fir, Abies grandis, A. grandis, oil, meal#

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claims 1, 2, 5, 6, 9, 12, 15-17, 26-46, 49, and 50, drawn to nucleotide sequences encoding myrcene synthase from Grand fir (Abies grandis), vectors comprising the nucleotide sequence, plants transformed with the vectors, seeds made by the transformed plants, and to a method of using the nucleotide sequence, the first product, the first method of making the product, and the first method of using the first product.

Group II, claims 1, 3, 5, 7, 10, 13, 15-17, 26-46, 49, and 50, drawn to nucleotide sequences encoding limonene synthase from Grand fir (Abies grandis), vectors comprising the nucleotide sequence, plants transformed with the vectors, and seeds made by the transformed plants, the second product.

Group III, claims 1, 4, 5, 8, 11, 14, 15-17, 26-46, 49, and 50, drawn to nucleotide sequences encoding pinene synthase from Grand fir (Abies grandis), vectors comprising the nucleotide sequence, plants transformed with the vectors, and seeds made by the transformed plants, the third product.

Group IV, claims 18-22 and 23, drawn to an isolated myrcene synthase protein from Grand fir (Abies grandis), the fourth product.

Group V, claims 18-22 and 24, drawn to an isolated limonene synthase protein from Grand fir (Abies grandis), the fifth product.

Group VI, claims 18-22 and 25, drawn to an isolated myrcene synthase protein from Grand fir (Abies grandis), the sixth product.

Group VII, claims 15-17, 26-46, 49, and 50, drawn to a method of using nucleotide sequences encoding limonene synthase from Grand fir (Abies grandis), the second method of using the second product.

Group VIII, claims 15-17, 26-46, 49, and 50, drawn to a method of using nucleotide sequences encoding pinene synthase from Grand fir (Abies grandis), the third method of using the third product.

Group IX, claims 47 and 48, drawn to oil and meal extracted from the seeds of Groups I-III, the seventh product.

The inventions listed as Groups I-IX do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the products of Groups I-VI and IX are distinct. The methods of Groups VII and VIII do not utilize the product of Group I. PCT Rule 13 does not provide for multiple methods of using within a single application (37 CFR 1.475(d)).

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/14528

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☒ No protest accompanied the payment of additional search fees.